WFABC: a Wright-Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data

Matthieu Foll[*,§,1,2], Hyunjin Shim[*,1,2], and Jeffrey D. Jensen[1,2]

[*] authors contributed equally

[§] corresponding author

[1]School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

[2]Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland

Address for all co-authors:

EPFL SV IBI-SV UPJENSEN

AAB 0 43 (Bâtiment AAB)

Station 15

CH-1015 Lausanne

SWITZERLAND

Keywords: Time-sampled data; Natural selection; Genetic drift; Effective population size; Population genetics; Approximate Bayesian Computation

Corresponding author: Matthieu Foll (matthieu.foll@epfl.ch)

Running title: ABC method for time-sampled data

1   Abstract

2   With novel developments in sequencing technologies, time-sampled data are

3   becoming more available and accessible. Naturally, there have been efforts in parallel to

4   infer population genetic parameters from these datasets. Here, we compare and analyze

5   four recent approaches based on the Wright-Fisher model for inferring selection

6   coefficients (s) given effective population size (N_e), with simulated temporal datasets.

7   Furthermore, we demonstrate the advantage of a recently proposed ABC-based method

8   that is able to correctly infer genome-wide average N_e from time-serial data, which is

9   then set as a prior for inferring per-site selection coefficients accurately and precisely.

10   We implement this ABC method in a new software and apply it to a classical time-serial

11   dataset of the *medionigra* genotype in the moth *Panaxia dominula*. We show that a

12   recessive lethal model is the best explanation for the observed variation in allele

13   frequency by implementing an estimator of the dominance ratio (h).

14   Introduction

15   The study of temporal changes in allele frequency originated with two of the

16   early founders of population genetics (Fisher 1922; Wright 1931), in which the fate of an

17   allele was considered under a variety of models – including neutrality, positive and

18   negative selection, and migration. Their celebrated debate on the relative roles of

19   selection and drift in shaping the course of evolution also encompassed time-sampled

20   data (Fisher and Ford 1947; Wright 1948), upon the publication of the time-series

21   analysis of the *medionigra* phenotype in the moth *Panaxia dominula*. Following on this,

22   and alternatively taking an experimental evolution approach, Clegg studied the

23   dynamics of gene frequency change in *Drosophila melanogaster* (Clegg *et al.* 1976;

24   Cavener and Clegg 1978; Clegg 1978). However, owing to the limited availability of

25   genetics markers, relatively few time-sampled datasets were available for consideration

26   throughout the remainder of the 20th century. Thus, most test statistics for

27   distinguishing the effects of selection from drift were focused on single time-point

28   datasets – basing inference on patterns in the site frequency spectrum, linkage

29   disequilibrium, and polymorphism/divergence (for a review, see Crisci *et al.* 2012).

30   Recently, sequencing data from multiple time-points has become increasingly

31   common owing to novel developments in sequencing technologies (Schuster 2007) –

32   coming from the fields of both ancient genomics and experimental evolution. This

33   additional temporal component has the promise of providing improved power for

34   inferring population genetic parameters compared to single time-point based analyses,

35   as the trajectory of the allele itself provides valuable information about the underlying

36   selection coefficient.

37   However, there are a limited number of methods currently available to estimate

38   these parameters from time-sampled data. Moment-based methods (Kimura and Crow

39   1963; Pamilo and Varvio-Aho 1980; Nei and Tajima 1981; Waples 1989; Jorde and

40   Ryman 2007) have been proposed utilizing the variance of gene frequency changes to

41   infer effective population size (N_e). In addition, likelihood-based methods (Williamson

42   and Slatkin 1999; Anderson *et al.* 2000; Berthier *et al.* 2002; Anderson 2005) have been

43   proposed to calculate the probability of a given data observation given a pre-defined

44   model. Efforts to incorporate selection into these estimation procedures have only

45   recently begun, and given the rapidly increasing availability of such sequencing datasets,

46   we now have a unique opportunity to re-address the puzzle of distinguishing genetic

47   drift from selection with greater precision and power.

3

48   Thus, we present here new software implementing and expanding an

49   approximate Bayesian computation (ABC, Sunnåker *et al.* 2013) approach to jointly infer

50   per-site effective population sizes (N_e) and selection coefficients (*s*) from time-sampled

51   data, initially described in Foll *et al.* (2014) to search for resistance mutations in time-

52   sampled data from the influenza virus. Furthermore, we compare this approach with

53   existing likelihood-based methods (Bollback *et al.* 2008; Malaspinas *et al.* 2012;

54   Mathieson and McVean 2013), in order to inform future users on the most suitable

55   method to be applied for any given dataset.

56   **Materials and Methods**

57   ***N_e-based ABC method***

58   The data *X* consists of allele frequency trajectories measured at *L* loci:

59   $x_i (i = 1,...,L)$. The N_e-based ABC methodology infers both the effective population size N_e

60   shared by all loci, and *L* locus-specific selection coefficients $s_i (i = 1,...,L)$. At a particular

61   locus *i*, we can approximate the joint posterior distribution as (see Foll *et al.* 2014 for

62   details):

$$P(N_e, s_i \mid X) \approx P(N_e \mid T(X)) P(s_i \mid N_e, U(X_i))$$

64   where $T(X) = T(X_1,...,X_L)$ denotes summary statistics that are a function of all loci

65   together chosen to be informative about N_e, and *U(X_i)* denotes locus-specific summary

66   statistics chosen to be informative about s_i. A two-step ABC algorithm as proposed by

67   Bazin *et al.* (2010) is used to approximate this posterior:

68   Step 1. Obtain an approximation of the density

$$P(N_e \mid T(X)) \approx P(N_e \mid X)$$

70   Step 2. For locus *i=1* to *i=L*:

71   i.   Simulate K trajectories $X_{i,k}$ from a Wright-Fisher model with s_i randomly

72   sampled from its prior and N_e from the density obtained in step 1.

73   ii.   Compute $U(X_{i,k})$ for each simulated trajectory.

74   iii.   Retain the simulations with the smallest Euclidian distance between $U(X_i)$

75   and $U(x_i)$ to obtain a sample from an approximation to

76            $$P(s_i \mid N_e, X_i)P(N_e \mid X) = P(N_e, s_i \mid X).$$

77   In the original algorithm (Bazin *et al.* 2010), the first step is also achieved using

78   ABC. In our case, we define $T(X)$ as a single statistic given by Jorde and Ryman (2007) *Fs'*

79   unbiased estimator of N_e:

80            $$Fs' = \frac{1}{t} \frac{Fs\left[1 - 1/(2\tilde{n})\right] - 2/\tilde{n}}{(1 + Fs/4)\left[1 - 1/(n_y)\right]} \text{ with } Fs = \frac{(x-y)^2}{z(1-z)}$$

81   where x and y are the minor allele frequencies at the two time points separated by t

82   generations, $z=(x+y)/2$, and $\tilde{n}$ is the harmonic mean of the sample sizes n_x and n_y at the

83   two time points expressed in number of chromosomes (twice the number of individuals

84   for diploids). We average *Fs'* values over sites and times to obtain a genome wide

85   estimator of $N_e = 1/Fs'$ for haploids and $N_e = 1/2Fs'$ for diploids (Jorde and Ryman 2007). A

86   Bayesian bootstrap approach (Rubin 1981) is used to obtain a distribution for

87    $P(N_e \mid T(X))$. Please note that we use the common notation where the effective

88   population size N_e corresponds to number of individuals, and the corresponding number

89   of chromosomes for diploids is $2N_e$.

90   In the second step, simulations are performed using a Wright-Fisher model with

91   an initial allele frequency and sample sizes matching the observed (simulation code

92   available in the downloadable software package). At each site, we utilize two summary

93   statistics derived from *Fs'*: $U(X_i) = (Fsd'_i, Fsi'_i)$ with *Fsd'* and *Fsi'* calculated respectively

94   between pairs of time points where the allele considered is decreasing and increasing in

95   frequency, such that at a given site $Fs'=Fsd' + Fsi'$. For the diploid model, we define the

96   relative fitness as $w_{AA}=1+s$, $w_{Aa}=1+sh$ and $w_{aa}=1$ where h denotes the dominance ratio (1

97   = dominant, 0.5 = co-dominance, 0 = recessive); and as $w_A=1+s$ and $w_a=1$ for the haploid

98   model (Ewens, 2004).

99   We here implement this two-step approach in a new command line C++ program

100   termed 'Wright-Fisher ABC' (WFABC). This estimation procedure is suitable for both

101   haploid and diploid models of selection. Source code and binary executables for Linux,

102   OS X, and Windows are freely available from the 'software' page of the Jensen Lab

103   website: http://jensenlab.epfl.ch/

104   **Likelihood-based methods**

105   Currently, there are three likelihood-based methods available for inferring

106   population genetic parameters from time-serial data (Bollback *et al.* 2008; Malaspinas *et*

107   *al.* 2012; Mathieson and McVean 2013), based on a Hidden Markov Model (HMM) to

108   model the allele frequency trajectory.

109   Bollback *et al.* (2008) co-estimates the selection coefficient (s) and the effective

110   population size (N_e) from a diffusion process, by approximating the Wright-Fisher model

111   and computing the maximum likelihood at fixed intervals. Malaspinas *et al.* (2012)

112   additionally estimates the allele age (t_0), and further approximates the Wright-Fisher

113   model through a one-step process. Mathieson and McVean (2013) estimates only the

114   selection coefficient (s) assuming that N_e is known, using an expectation-maximization

115   (EM) algorithm that can be extended to the case of a structured population. The fitness

116   is parameterized as with our N_e-based ABC approach (see above). The likelihood

117   function of the parameters of interest - $\theta=(\gamma,N_e)$ for the Bollback *et al.* (2008) method

118   and $\theta=(\gamma,N_e,t_0)$ for the Malaspinas *et al.* (2012) method, where $\gamma=2N_e s$ – is conditioned

119   over all population allelic frequencies $x_{j_1},...,x_{j_m}$ at sampling times $T=(t_1,...,t_m)$, and is

120   given as:

121   $$l(\theta) = \sum_{j_1}...\sum_{j_m} p(i_1,...,i_m \mid \theta,T,x_{j_1},...,x_{j_m})p(x_{j_1},...,x_{j_m} \mid \theta,T)$$

122   where i_k is the frequency of the minor allele at the sampling time k, and j_k is the true

123   minor allele frequency at the sampling time k, where $x_{j_k} = j_k / 2N_e$. The first term of the

124   likelihood is the emission probability, which is modeled as a binomial sampling, and the

125   second term represents the transition probabilities in the Markov chain. In both

126   methods the Markov chain is approximated by a diffusion process, from which the

127   transition probabilities are given as the backward Kolmogorov equations (Ewens, 2004).

128   The major difference between these three likelihood-based methods comes from

129   the implementation of how these probabilities are calculated. The Bollback *et al.* (2008)

130   method utilizes numerical approximations to evaluate the likelihood function, first by

131   using the Crank-Nicolson approximation (Crank *et al.* 1947) for the backward

132   Kolmogorov equation, and second by using numerical integration for the emission

133   probability. Mathiesion and McVean (2013) use an expectation-maximization (EM)

134   algorithm to find the maximum likelihood estimate (MLE) of *s* based on the MLE for

135   complete observations (*i.e.* at every generations). Malaspinas *et al.* (2012) approximates

136   the diffusion process by a one-step process. The state space of the process is the

137   population allele frequencies that are denoted by $(z_0,...,z_{H-1})$, where $z_0=0$, $z_{H-1}=1$ and $z_{k-1}<z_k$.

138   $_1<z_k$. The one-step process only allows transitions between two adjacent states (i.e., from

139   z_i to z_{i-1}, or z_i to z_{i+1}, hence the infinitesimal generator Q can be constructed as a

140   tridiagonal *HxH* matrix:

141
$$Q = \begin{pmatrix} 0 & 0 & 0 & & \cdots & & 0 \\ \delta_1 & \eta_1 & \beta_1 & & & & \\ 0 & \ddots & \ddots & \ddots & & & \vdots \\ & & \delta_i & \eta_i & \beta_i & & \\ \vdots & & & \ddots & \ddots & \ddots & 0 \\ & & & & \delta_{H-2} & \eta_{H-2} & \beta_{H-2} \\ 0 & & \cdots & & 0 & 0 & 0 \end{pmatrix}$$

142   where β_i denotes the transition rate from z_i to z_{i+1}, δ_i the transition rate from to z_i to z_{i-1},

143   and η_i is the rate of no transition such that $\eta_i=1-(\beta_i+\delta_i)$. The appropriate choice for the

144   parameters β and δ of the matrix Q are given for both the diploid and the haploid

145   Wright-Fisher model in Malaspinas *et al.* (2012) and Foll *et al.* (2014), respectively.

146   **Simulated datasets for testing**

147   For real data, it is important to take into account the non-random criteria one

148   used to select sites from the genome for analyses. This so-called ascertainment bias is

149   known to be very important for single time point SNP data (Nielsen and Signorovitch

150   2003), but has not been studied so far for time-sampled data. One of the reasons is that

151   including realistic ascertainment schemes in likelihood-based methods is a difficult task.

152   The one-step process used in Malaspinas *et al.* (2012) can be adjusted in order to match

153   the way in which ancient DNA data is generally collected, such that the locus considered

154   is polymorphic at the present time. This condition implies that the process can never

155   reach the absorbing states 0 and 1, and one needs to remove the first and last rows and

156   columns of the Q matrix. In the current implementation of Malaspinas *et al.* (2012) only

157   this conditional case is available and has been tested for this study. The Bollback *et al.*

158   (2008) method implements an unconditional model (no ascertainment) as well as the

159 particular case where the allele is known to be beneficial and reaches fixation during the

160 sampling period. There is no ascertainment model implemented in Mathieson and

161 McVean (2013) method. One distinct advantage of this simulation-based approach is the

162 ability to easily incorporate different ascertainment schemes into the estimation

163 procedure, as one simply needs to be able to simulate them. We here present three such

164 non-exclusive schemes: (i) observing a minimum allele frequency over the entire

165 trajectory, (ii) observing a minimum allele frequency at the last time point (including

166 fixation like in the Bollback *et al.* (2008) method), and (iii) being polymorphic at the last

167 time point (like in the Malaspinas *et al.* (2012) method). We note in particular that the

168 first case is something that will be present in any data set but is not available in

169 likelihood-based methods. Because the three available likelihood methods implement

170 different ascertainment processes, and these processes lead to more or less informative

171 data, it is not possible to make a direct comparison of their performance. For this reason,

172 we separately compare them with WFABC.

173   For this comparative study, we generate simulated datasets using the Wright-

174 Fisher model with a range of parameter values for the effective population size (N_e) and

175 selection coefficient (s). For the diploid Wright-Fisher model, co-dominant time-serial

176 allele frequency data from 1000 replicates for N_e=(200, 1000, 5000) and $s \in [0,0.4]$ are

177 generated. In order to assess the precision and accuracy of these methodologies in

178 estimating two potential empirical cases of small s values and large s values, the

179 selection coefficients are divided into two sets of s=(0, 0.005, 0.01, 0.015, 0.02) and s=(0,

180 0.1, 0.2, 0.3, 0.4). For WFABC, we retained the best 1% of 500'000 simulations based on

181 the Euclidian distance between the observed and simulated *Fsd'* and *Fsi'* statistics and

182 use the mean of the posterior distributions obtained for s using a rejection ABC

9

183   algorithm (Sunnåker *et al.* 2013) as a point estimate. Firstly, we use these simulated

184   datasets to demonstrate the performance of WFABC when different sampling time

185   points and different sample sizes are used. Secondly, we show the influence of the

186   ascertainment procedure using two examples: an unconditional but unrealistic case

187   where all trajectories start with an initial minor allele frequency of 10%, and one

188   ascertained case where a new mutation occurs at the first generation and only the

189   trajectories reaching a frequency of 5% at least in one sampling time point are kept. We

190   use these simulated datasets to compare the performance of WFABC with the method of

191   Mathieson and McVean (2013).

192   Finally, we focus on the ascertainment case of the allele segregating at the last

193   sampling time point, as this model is the only realistic one allowing us to compare

194   WFABC with a likelihood method. Depending on the strength of selection and the

195   effective population size, mutations reach fixation more or less rapidly. In order to

196   generate data with the condition of being polymorphic at the last sampling time point,

197   the number of generations is adjusted to have a non-zero probability for this condition,

198   allowing us to efficiently simulate such scenarios. A new mutation occurs at the first

199   generation and 100 samples are drawn randomly through binomial sampling, with 12

200   sampling time points. Using these simulated datasets, comparative studies of WFABC

201   with Malaspinas *et al.* (2012) and Bollback *et al.* (2008) methods are carried out, with

202   the search range and the prior for the selection coefficient set as $s \in [-0.1, 0.1]$ for small

203   values ranging from 0 to 0.02, and $s \in [-0.2, 0.6]$ for large values ranging from 0 to 0.4.

204   Simulated datasets from the haploid Wright-Fisher model are also generated, but only

205   for one set of $N_e = 1000$ in order to validate the performance of the modified haploid

206   version of Malaspinas *et al.* (2012) model described in Foll *et al.* (2014). For Malaspinas

10

207   *et al.* (2012), we used the quadratic grid option (Gutenkunst *et al.* 2009) for computing

208   the likelihood and the Nelder-Mead simplex algorithm option to find the maximum

209   likelihood (Nelder and Mead 1965).

210   Both the Bollback *et al.* (2008) and Malaspinas *et al.* (2012) methods estimate

211   parameters based on a single allele frequency trajectory, which thus contains limited

212   information about N_e. WFABC utilizes multiple sites in order to estimate N_e, which is

213   then used as a prior to estimate selection coefficients. In order to implement an equal

214   comparison, we fix N_e to its true value in all scenarios and evaluate the ability of these

215   approaches to estimate selection coefficients.

216   In order to demonstrate the advantage of estimating N_e correctly, a final multi-

217   locus scenario is generated where both s and N_e values are inferred. For N_e fixed at 1000,

218   10'000 trajectories are simulated with 500 being under selection with s randomly

219   chosen in $[0.05, 0.4]$. We used a search range of $\gamma \in [0, 2000]$ and $N_e \in [50, 2000]$ for the

220   Malaspinas *et al.* (2012) method, and a uniform prior $s \in [-0.2, 0.6]$ for WFABC. For

221   WFABC, all the 10'000 trajectories are used in the first step to obtain a posterior

222   distribution for N_e, which is used in the second step to estimate s at each locus

223   individually as explained above.

224   **Results**

225   **Performance of the examined estimation procedures**

226   The performance studies of WFABC are presented in a standard boxplot with the

227   box as the first, second, and the third quartile, and the whiskers as the lowest and

228   highest datum within the 1.5 interquartile range of the lower and upper quartile,

229   respectively. The first boxplot shows the estimated selection coefficients for different

230 numbers of sampling time points as 12, 6 or 2 (Figure S1). As expected, a larger number

231 of sampling time points yields a better estimate of s – however, WFABC is able to

232 estimate s accurately with as small as 2 sampling time points for moderate values of

233 *s<0.2*. The second boxplots shows the estimated selection coefficients from WFABC for

234 the sample sizes of 1000, 100 and 20 with N_e=1000 (Figure S2). The estimation of s

235 improves as the number of sample sizes increases as expected.

236   For the comparative studies of WFABC with the Mathieson and McVean (2013)

237 method, the unconditional case with an initial minor allele frequency of 10% is shown in

238 boxplots for the small s values (Figure S3) and the large s values (Figure S4). For the

239 small s values, the Mathieson and McVean (2013) method performs better than WFABC.

240 However, for larger values of s, the Mathieson and McVean (2013) method shows an

241 increasing trend of underestimation, whereas WFABC remains unbiased (Figure S4). For

242 the conditional case of ascertaining the simulated datasets with a minimum frequency of

243 5% at one sampling time point, the performance of WFABC is noticeably better for both

244 the small and large s values (Figure 1, Figure 2). Figure 1 shows that the Mathieson and

245 McVean (2013) method constantly overestimates small s values. Figure 2 indicates that

246 this biased is compensated by the underestimation shown above for large s values.

247   For the conditional case studies of the allele segregating at the last sampling time

248 point, we obtain 6 sets of results with the varying parameters for the diploid model and

249 2 sets of results for the haploid model from the three approaches. Results are given for

250 N_e = 1000 for small s values (Figure 3), and large s values (Figure 4). Additionally, tables

251 with the calculations of the root mean square error (RMSE) and bias for each set of

252 results are shown for the small s values and the large s values in Table 1 and Table 2,

253 respectively. Note that MSE is defined as the sum of the variance and the squared bias of

12

254 the estimator, and therefore incorporates information from both precision (variance)

255 and accuracy (bias).

256   Comparing the three approaches for small s values (Figure 3), WFABC and the

257 Malaspinas $et~al.$ (2012) approach produce good estimates of s, whereas the Bollback et

258 $al.$ (2008) method fails to infer different s values. From the boxplot, both the median and

259 the mean of WFABC and the Malaspinas $et~al.$ (2012) approach are close to the true s

260 value. However, WFABC appears to contain a longer tail of under-estimated s values, and

261 the interquartile range boxes are wider. Table 1 provides more quantitative

262 comparisons of their performance. The RMSE values reveal that for all the cases of

263 N_e=1000 and small s, the Malaspinas $et~al.$ (2012) approach is generating more precise

264 estimates. On the other hand, WFABC produces estimates of less bias for all small s

265 values in this set.

266   Comparing the three approaches for large s values (Figure 4), WFABC and the

267 Malaspinas $et~al.$ (2012) approach produce reasonable estimates of s, whereas the

268 Bollback $et~al.$ (2008) method again fails to detect any difference in s values. For s values

269 larger than 0.1, the performance of WFABC is significantly better than the Malaspinas et

270 $al.$ (2012) approach both in accuracy and precision (Table 2) – producing estimates with

271 10-fold less bias and 2-fold less error than the Malaspinas $et~al.$ (2012) approach. This

272 gap in performance increases from $s = 0.2$ to $s = 0.4$, thus this trend may be extrapolated

273 to higher s values.

274   Notably, the Bollback $et~al.$ (2008) method consistently estimates $s = 0$ for all

275 examined datasets. This poor performance has been evaluated for various conditions

276 including changing grid sizes, conditioning on fixation, and varying sampling time points

277 – with no perceptible difference in results. In order to verify our usage, the method was

278   tested with the exact parameters utilized in the initial paper for the example of

279   bacteriophage MS2 (Bollback *et al.* 2008), resulting in the successful replication of their

280   results (Figure S5, Figure S6). Notably, the performance of the statistic depends on the

281   choice of the search range for γ, due to the presence of local peaks in the likelihood

282   function. When the interval is chosen to be narrow and centered around the true *s* value

283   of 0.4, the estimated *s* is correctly given owing to the local maximum (replicating their

284   result). However, when the full likelihood surface is examined, the global maximum is

285   present near 0 as observed in all simulated test replicates. For this reason, the Bollback

286   *et al.* (2008) method is excluded from the further analyses of performance, as well as

287   from the illustrative data application.

288   As shown, our comparison studies suggest that WFABC and Malaspinas *et al.*

289   (2012) approaches perform almost equally well for estimating selection coefficients for

290   both small and large values. While WFABC slightly overestimates the selection

291   coefficient in the cases of small *s* and small N_e (Figure S7, Table S1), it is notable that the

292   Malaspinas *et al.* (2012) performs particularly well when *s* is in the range of 0.01 and

293   0.02, as observed by Malaspinas *et al.* (2012). In contrast, WFABC exhibits less bias

294   when *s* values are large (i.e., greater than 0.1) for small N_e (Figure S8, Table S2). Thus,

295   we conclude that the two methods estimate the selection coefficient to a high accuracy

296   and precision for $N_e \in [200,1000]$ and for the condition of the allele segregating at the

297   last sampling time point. In general, although the difference in performance both in

298   precision and accuracy is minor, the Malaspinas *et al.* (2012) approach appears to give

299   superior results for small *s* values, whereas WFABC for large *s* values.

300   However, the Malaspinas *et al.* (2012) approach encounters a limitation in

301   computation efficiency for the set of N_e=5000 (Figures S9, S10; Tables S3, S4). For the

302   large s values, the computation time was too lengthy to complete the 1000 replicates

303   (see below), thus only the results from WFABC are shown (Figure S10). Compared to the

304   small N_e values, the estimation of large s is becoming more accurate and precise as N_e

305   gets large (Tables 2, S2, S4). This trend is the same for small s values (Tables 1, S1, S3),

306   although the bias appears to switch from over-estimation to under-estimation as N_e

307   increases. For the Malaspinas et $al.$ (2012) approach, the accuracy of inferring small s

308   values improves as N_e increases, although the precision remains similar (Tables 1, S1,

309   S3). However, for N_e =5000 and s>0.01, the 1000 replicates were not complete due to

310   the lengthy computational time, thus the RMSE and bias values are not available (Table

311   S3).

312   For the haploid model, WFABC shows superiority in both accuracy and precision

313   for inferring any s values compared with the Malaspinas et $al.$ (2012) approach (Figures

314   S11 and S12, Tables S5 andS6).

315   Finally, the multi-locus scenario demonstrates the great benefit provided by the

316   ability of WFABC to use the information shared by all loci to estimate N_e. The RMSE of

317   the selection coefficients calculated over the 500 trajectories under selection is less than

318   half that obtained using the Malaspinas et $al.$ (2012) approach (0.049 vs. 0.10, see Figure

319   5).

320   In summary, WFABC is superior for diploid cases when both s and N_e values are

321   large (i.e., $\gamma=2N_es$ is large), for any haploid cases, and when multiple loci are available,

322   whereas the Malaspinas et $al.$ (2012) approach is suitable for cases when γ values are

323   small.

15

Comparison of computational efficiency

324

325   Apart from accuracy and precision, an important difference in performance

326 between these methods is the computational efficiency. A major advantage of WFABC is

327 the computational speed, which, for example, allows for an evaluation of all observed

328 sites in the genome in order to identify putatively selected outliers (Foll *et al.* 2014), and

329 to estimate the full distribution of fitness effects of segregating mutations. For the

330 Malaspinas *et al.* (2012) approach, the computational time becomes heavy when γ is

331 larger than 200 and is no longer feasible when γ is approaching 1000, whereas WFABC

332 has no restriction on the sizes of N_e and s. For the Mathieson and McVean (2013)

333 approach, the CPU time of estimating only the selection coefficient of each site is around

334 2 seconds regardless of the size of N_e, but still is slower than WFABC. Therefore, we

335 suggest that the likelihood-based approach is preferable in cases where both the

336 candidate mutation and effective population sizes are known *a priori*, whereas WFABC

337 is preferable in the absence of this information. The average CPU time spent for each

338 replicate for the diploid model is shown in Table 3. We note that when the Malaspinas *et*

339 *al.* (2012) method is also used to estimate N_e, the difference in CPU time between the

340 two methods is even greater.

Data Application

341

342   We applied WFABC to a time-serial data set of the *medionigra* morph in a

343 population of *Panaxia dominula* (scarlet tiger moth) at Cothill Fen near Oxford. This

344 colony was first studied by Fisher and Ford (1947), and further observations have been

345 collected almost every year until at least 1999 (Cook and Jones 1996; Jones 2000). The

346 moth *P. dominula* has a one year generation time and lives near the Oxford district in

347 isolated colonies. The typical phenotype has a black fore wing with white spots and a

348 scarlet hind wing with black patterns (see Fisher and Ford 1947). The *medionigra* allele

349 produces the *medionigra* phenotype when heterozygous, and the *bimacula* phenotype

350 when homozygous, changing the pigment and patterns on the wings to an increasing

351 degree, and is almost never observed (Sheppard and Cook 1962). Using our notation

352 above, we denote by *A* the *medionigra* allele and the fitness of the three genotypes are

353 given by $w_{AA}=1+s$ (*bimacula*), $w_{Aa}=1+sh$ (*medionigra*) and $w_{aa}=1$.

354   The respective role of drift and natural selection to explain the rapid decline of

355 the *medionigra* allele frequency after 1940 (Figure 6) was the subject of a strong debate

356 between Fisher and Wright (Fisher and Ford 1947; Wright 1948), with Wright arguing

357 that the observed pattern until 1946 could be explained by drift alone with an effective

358 population size of N_e=150 (Wright 1948). The same data containing further

359 observations has been re-analyzed several times (Cook and Jones 1996; O'Hara 2005;

360 Mathieson and McVean 2013) with most studies concluding that the *medionigra* allele is

361 negatively selected with s=-0.14 (Cook and Jones 1996) or s=-0.11 (Mathieson and

362 McVean 2013) based on a co-dominant model (*h=1/2*). Recently, Mathieson and McVean

363 (2013) found that a fully recessive *medionigra* allele (*h=0*) fits with a higher likelihood

364 compared to *h=1/2* but with a much larger selection coefficient *s≈-1*. In particular, this

365 recessive lethal model explains better the persistence of the *medionigra* allele at a low

366 frequency for so many generations (Mathieson and McVean 2013). However this large

367 value of *s* is outside the range for which their approximations are valid and this

368 hypothesis could not be formally tested.

369   Our ABC approach based on simulations can deal with the full range of *s* values

370 and we further extended it here to co-estimate the degree of dominance *h*. We followed

17

371   the intuitive idea of Mathieson and McVean (2013) that the number of generations

372   during which the allele persists at low frequency is informative for the degree of

373   dominance *h.* More formally, we added two summary statistics in our ABC procedure, *tl*,

374   defined as the number of generations where the allele frequency is below 5% and not

375   lost; and *th* defined as the number of generations where the allele frequency is above

376   95% and not fixed. For the moth data, we have *tl=54* and *th=0* (see Figure 6). Using the

377   simulated distributions, *Fsd'* and *Fsi'* are both normalized by the largest standard

378   deviation *max(sd(Fsd'),sd(Fsi'))*, as well as *tl* and *th* by *max(sd(tl),sd(th))*. We followed

379   Mathieson and McVean (2013) and ran our ABC method using a fixed population size of

380   *$2N_e=1000$* and we plot the corresponding joint posterior distribution for *s* and *h* in

381   Figure 7A. The mode of the joint posterior distribution is at *s=-1* and *h=0.043*, supporting

382   the idea of a lethal *bimacula* phenotype (*$w_{AA}=0$*) with a deleterious *medionigra*

383   phenotype (*$w_{Aa}=0.96$*) consistent with previous observations (Sheppard and Cook 1962).

384   The shape of the joint posterior distribution (Figure 7A) shows that the *medionigra*

385   allele is either very strongly selected against and almost completely recessive (*h≈0*), or

386   co-dominant with a weaker selection coefficient (*s≈-0.2*). Even if the density is larger in

387   the recessive lethal region of the parameter space (bottom left corner in Figure 7A), the

388   two-dimensional 90% highest posterior region includes the alternative co-dominant

389   hypothesis. As it has been argued that the effective population size could be of the order

390   of a few hundred (Wright 1948; O'Hara 2005), we also ran the analysis with *$2N_e=100$*.

391   The joint posterior distribution for *s* and *h* in Figure 7B shows a similar pattern with a

392   mode at *s=-1* and *h=0*. However the surface is flatter and the two-dimensional 90%

393   highest posterior region now includes *s=0,* confirming Wright's view that a small enough

394   population size could explain the observed pattern with genetic drift alone (Wright

18

395   1948; Mathieson and McVean 2013). We finally ran the analysis using a uniform prior

396   for $2N_e$ between 100 and 10000 to take into account its uncertainty in the estimate. In

397   this case the joint posterior gives a stronger support for a lethal *bimacula* phenotype

398   with a deleterious *medionigra* phenotype as compared to $2N_e=1000$ (Figure S13).

Discussion

400   Maximum likelihood estimators (MLE) have the advantage of being consistent

401   and efficient, but the computational method used to find the maximum likelihood can be

402   critical in complex models. For instance, even though the underlying model of the

403   Bollback *et al.* (2008) approach is similar to the Malaspinas *et al.* (2012) approach and

404   the Mathieson and McVean (2013) approach, the difference in performance appears to

405   be coming from the difference in the implementation of computational methods. On the

406   other hand, ABC-based methods reduce datasets into summary statistics, and thus the

407   performance of these methods is dependent on the chosen statistics. The difference in

408   performance demonstrated through this comparative study between our newly

409   proposed ABC-based method and the likelihood-based methods is very small in most

410   cases. Therefore, we can hypothesize that the two summary statistics - *Fsd'* and *Fsi'* - are

411   close to being statistically sufficient.

412   A disadvantage of the likelihood methods arises from the limitations imposed by

413   the assumptions made for diffusion approximation. In order to approximate the Wright-

414   Fisher model with the diffusion process, one makes the assumption that the Markov

415   process is continuous in state space and time. This assumption requires s to be small and

416   N_e to be large (Durrett, 2008). Thus, likelihood methods based on diffusion

417   approximation are inevitably limited to the cases of large effective population sizes and

19

418   small selection coefficients, and may explain the biases observed at large s values in this

419   study. Additionally the computational efficiency limits the value of γ to be under 1000

420   for the Malaspinas *et al.* (2012) approach, even if s is small, N_e is again limited to be

421   under 5000. Thus, despite having good performance, the likelihood methods are limited

422   to cases of small s and intermediate N_e values.

423   Although WFABC gives slightly less precise results in some cases of small s, the

424   two major advantages of this implementation comes from its ability to consider complex

425   ascertainment cases and to accurately estimate N_e from multi-locus genome wide data.

426   The Mathieson and McVean (2013) method is computationally efficient but is based on

427   the unrealistic assumptions that N_e is known and that the data is completely

428   unascertained. Using a point estimate of N_e obtained from the first step of WFABC in the

429   Mathieson and McVean (2013) method would ignore the uncertainty on this parameter.

430   Malaspinas *et al.* (2012) only handles the conditional cases of polymorphism at the last

431   sampling time point and estimates N_e separately at each site. WFABC is computational

432   efficient and enables the estimation of a genome-wide average N_e from time-sampled

433   data, after which per-site selection coefficients may be estimated efficiently and

434   accurately. There currently exist no such likelihood-based methods to estimate N_e

435   accurately from time-sampled genomic data, since the information is only based on a

436   single allele trajectory. Our simulated multi-locus scenario shows that estimating N_e

437   accurately leads to a higher precision in the estimates of selection coefficients s, leading

438   overall to a smaller RMSE as compared to the Malaspinas *et al.* (2012) method. As the

439   effective population size is an important population genetic parameter that is often

440   unknown in both ancient genomics and experimental evolution, WFABC provides a

441   practical and flexible platform to be utilized in any time-serial data for efficiently

442   inferring a wide range of N_e and s values to high accuracy. Finally, we note that the ABC

443   method also has the advantage of providing a posterior distribution rather than simply a

444   point estimate, allowing for easy-to-build credibility intervals.

445   The application of WFABC to the *P. dominula* data confirms that a nearly fully

446   recessive lethal model for the *medionigra* allele is the best explanation for the observed

447   pattern as hypothesized by Mathieson and McVean (2013). We note that in this case,

448   once the allele frequency is low enough such that heterozygotes almost never occur, it

449   behaves like a neutral model. We used this feature to estimate N_e using *Fs'* (Jorde and

450   Ryman 2007) by considering only time-points after 1950, when the *medionigra* allele

451   first reaches a frequency below $\sqrt{0.001} \approx 0.032$ and we obtained *2N$_e$=927*, which is

452   consistent with previous estimates (Fisher and Ford 1947; Cook and Jones 1996; O'Hara

453   2005). This application also demonstrates that WFABC is very flexible, as it can also be

454   used to co-estimate *h* and accommodate very large selection coefficients (such as *s=-1*, as

455   in our application here). In order to confirm the validity of our approach in this case, we

456   simulated 1000 datasets mimicking the *P. dominula* data (same number of generations,

457   time points, sample sizes and initial allele frequency). We fixed *s=-1* and *h=0.05* and let

458   *2N$_e$* vary uniformly between 100 and 10000 for each simulation, and estimated *s* and *h*

459   using our ABC method (Figure 8). Both parameter estimates are unbiased and while

460   distinguishing lethality (*s=-1*) from very strong negative selection (*s<-0.5*) seems to be

461   difficult, *h* is estimated with a very small variance.

462   Finally, it should be noted that all the methods described and utilized in this

463   study assume that the loci are in linkage equilibrium and take no demographic history

464   into account except the Mathieson and McVean (2013) method which integrates

465   structured populations with a lattice model. Owing to the generally low number of

466   generations in time series data, one expects only little recombination to occur and long

467   stretched of DNA may be under linkage disequilibrium. Foll *et al.* (2014) demonstrated

468   that the model is robust to fluctuating population sizes, but this may not hold for all

469   demographic scenarios. It is an important future challenge to expand these methods

470   further for inferring demographic parameters from time-serial data. WFABC has great

471   potential in achieving this challenging task, as ABC-based estimators lend themselves

472   more readily to the incorporation of complex demographic models compared to

473   likelihood-based methods.

474   **Acknowledgements**

479

References

480   **References**

481   Anderson EC(2005) An Efficient Monte Carlo Method for Estimating Ne From

482   Temporally Spaced Samples Using a Coalescent-Based Likelihood. *Genetics,* **170,** 955-

483   967.

484   Anderson EC, EG Williamson, EA Thompson (2000) Monte Carlo evaluation of the

485   likelihood for Ne from temporally spaced samples. *Genetics,* **156,** 2109-2118.

486   Bazin E, KJ Dawson, MA Beaumont (2010) Likelihood-free inference of population

487   structure and local adaptation in a Bayesian hierarchical model. *Genetics,* **185,** 587-602.

488   Berthier P, MA Beaumont, J-M Cornuet, G Luikart (2002) Likelihood-based estimation of

489   the effective population size using temporal changes in allele frequencies: a genealogical

490   approach. *Genetics,* **160,** 741-751.

491   Bollback JP, TL York, R Nielsen (2008) Estimation of 2Nes from temporal allele

492   frequency data. *Genetics,* **179,** 497-502.

493   Cavener DR, MT Clegg (1978) Dynamics of correlated genetic systems. IV. Multilocus

494   effects of ethanol stress environments. *Genetics,* **90,** 629-644.

495   Clegg MT(1978) Dynamics of correlated genetic systems. II. Simulation studies of

496   chromosomal segments under selection. *Theor Popul Biol,* **13,** 1-23.

497   Clegg MT, JF Kidwell, MG Kidwell, NJ Daniel (1976) Dynamics of correlated genetic

498   systems. I. Selection in the region of the Glued locus of Drosophila melanogaster.

499   *Genetics,* **83,** 793-810.

500   Cook LM, DA Jones (1996) The medionigra Gene in the Moth Panaxia dominula: The

501   Case for Selection. *Philosophical Transactions: Biological Sciences,* **351,** 1623-1634.

502   Crank J, P Nicolson, P.Nicolson (1947) A practical method for numerical evaluation of

503   solutions of partial differential equations of the heat-conduction type. *Mathematical*

504   *Proceedings of the Cambridge Philosophical Society,* **43,** 50-67.

505   Crisci JL, Y-P Poh, A Bean, A Simkin, JD Jensen (2012) Recent progress in polymorphism-

506   based population genetic inference. *J Hered,* **103,** 287-296.

507   Durrett, R., 2008 *Probability models for DNA sequence evolution.* Springer, New York.

508   Ewens, W. J., 2004 *Mathematical population genetics : theoretical introduction.* Springer,

509   New York.

510   Fisher RA(1922) On the dominance ratio. *Proceedings of the royal society of Edinburgh,*

511   **42,** 321-341.

512   Fisher RA, EB Ford (1947) The spread of a gene in natural conditions in a colony of the

513   moth Panaxia dominula L. *Heredity (Edinb),* **1,** 143-174.

514   Foll M, Y-P Poh, N Renzette, A Ferrer-Admetlla, *et al.* (2014) Influenza Virus Drug

515   Resistance: A Time-Sampled Population Genetics Perspective. *PLoS Genet,* **10,** e1004185.

516   Gutenkunst RN, RD Hernandez, SH Williamson, CD Bustamante (2009) Inferring the

517   joint demographic history of multiple populations from multidimensional SNP

518   frequency data. *PLoS Genet,* **5,** e1000695.

519   Jones DA(2000) Temperatures in the cothill habitat of Panaxia (Callimorpha) dominula

520   L. (the scarlet tiger moth). *Heredity (Edinb),* **84 (Pt 5),** 578-586.

521   Jorde PE, N Ryman (2007) Unbiased estimator for genetic drift and effective population

522   size. *Genetics,* **177,** 927-935.

523   Kimura M, JF Crow (1963) The Measurement of Effective Population Number. *Evolution,*

524   **17,** 279-288.

525   Malaspinas A-S, O Malaspinas, SN Evans, M Slatkin (2012) Estimating allele age and

526   selection coefficient from time-serial data. *Genetics,* **192,** 599-607.

527   Mathieson I, G McVean (2013) Estimating selection coefficients in spatially structured

528   populations from time series data of allele frequencies. *Genetics,* **193,** 973-984.

529   Nei M, F Tajima (1981) Genetic drift and estimation of effective population size. *Genetics,*

530   **98,** 625-640.

531   Nelder JA, R Mead (1965) A simplex method for function minimization. *The computer*

532   *journal,* **7,** 308-313.

533   Nielsen R, J Signorovitch (2003) Correcting for ascertainment biases when analyzing

534   SNP data: applications to the estimation of linkage disequilibrium. *Theor Popul Biol,* **63,**

535   245-255.

536   O'Hara RB(2005) Comparing the effects of genetic drift and fluctuating selection on

537   genotype frequency changes in the scarlet tiger moth. *Proceedings of the Royal Society of*

538   *London Series B-Biological Sciences,* **272,** 211-217.

539   Pamilo P, SL Varvio-Aho (1980) On the estimation of population size from allele

540   frequency changes. *Genetics,* **95,** 1055-1057.

541   Rubin DB(1981) The Bayesian Bootstrap. *The Annals of Statistics,* **9,** 130-134.

542   Schuster SC(2007) Next-generation sequencing transforms todays biology. *Nature,* **200,** .

543   Sheppard PM, LM Cook (1962) The manifold effects of the Medionigra gene of the moth

544   Panaxia dominula and the maintenance of a polymorphism. *Heredity (Edinb),* **17,** 415-

545   426.

546   Sunnåker M, S Wodak, AG Busetto, E Numminen, *et al.* (2013) Approximate Bayesian

547   Computation. *PLoS Comput Biol,* **9,** e1002803.

548   Waples RS(1989) A generalized approach for estimating effective population size from

549   temporal changes in allele frequency. *Genetics,* **121,** 379-391.

550   Williamson EG, M Slatkin (1999) Using maximum likelihood to estimate population size

551   from temporal changes in allele frequencies. *Genetics,* **152,** 755-761.

552   Wright S(1931) Evolution in Mendelian Populations. *Genetics,* **16,** 97-159.

553   Wright S(1948) On the Roles of Directed and Random Changes in Gene Frequency in the

554   Genetics of Populations. *Evolution,* **2,** 279-294.

Data Accessibility

555   **Data Accessibility**

556   This study is primarily based on simulated data created using the WFABC software

557   available from the "software" page at http://jensenlab.epfl.ch/. The *P. dominula* moth

558   data set has been taken from (Cook and Jones 1996; Jones 2000) and is also provided in

559   the WFABC package.

560   **Author Contributions**

561   M.F. and J.D.J conceived the idea. M.F. developed the WFABC software and H.S. extended

562   the Malaspinas et al. (2012) software for haploids. M.F. performed the simulations and

563   the WFABC analyses. H.S. performed the Mathieson and McVean (2013), Bollback *et al.*

564   (2008) and Malaspinas *et al.* (2012) analyses. All authors contributed to writing the

565   manuscript.

Tables

Table 1. RMSE and bias for the small s and N_e=1000 scenario for the Wright-Fisher diploid model

		s=0	s=0.005	s=0.01	s=0.015	s=0.02
Bias	WFABC	-0.0035	-0.0046	-0.0032	-0.0014	-0.0015
	Malaspinas *et al.* (2012)	-0.0044	-0.0050	-0.0039	-0.0026	-0.0022
RMSE	WFABC	0.017	0.018	0.017	0.016	0.018
	Malaspinas *et al.* (2012)	0.013	0.014	0.012	0.011	0.011

Table 2. RMSE and bias for the big s and N_e=1000 scenario for the Wright-Fisher diploid model

		s=0	s=0.1	s=0.2	s=0.3	s=0.4
Bias	WFABC	0.017	-0.017	0.0064	-0.0044	-0.0083
	Malaspinas *et al.* (2012)	-0.0024	-0.018	-0.025	-0.056	-0.091
RMSE	WFABC	0.059	0.060	0.050	0.046	0.059
	Malaspinas *et al.* (2012)	0.047	0.046	0.048	0.084	0.13

Table 3. Average CPU time consumed in seconds to run one simulation replicate.

	N_e=200		N_e=1000		N_e=5000	
	Small s	Large s	Small s	Large s	Small s	Large s
WFABC	5	3	15	6	300	70
Malaspinas *et al.* (2012)	380	250	350	18000	80000	-

Figure 1. Boxplot for the estimated small selection coefficients from each simulated replicate of the Wright-Fisher diploid model, with N_e=1000, 90 generations, and the minimum frequency of 5% at one sampling time point. The gray rectangles correspond to WFABC and the white rectangles to the Mathieson and McVean (2013) method. The red circles are the true values for *s*, and the blue triangles are the mean of the estimated *s* values.
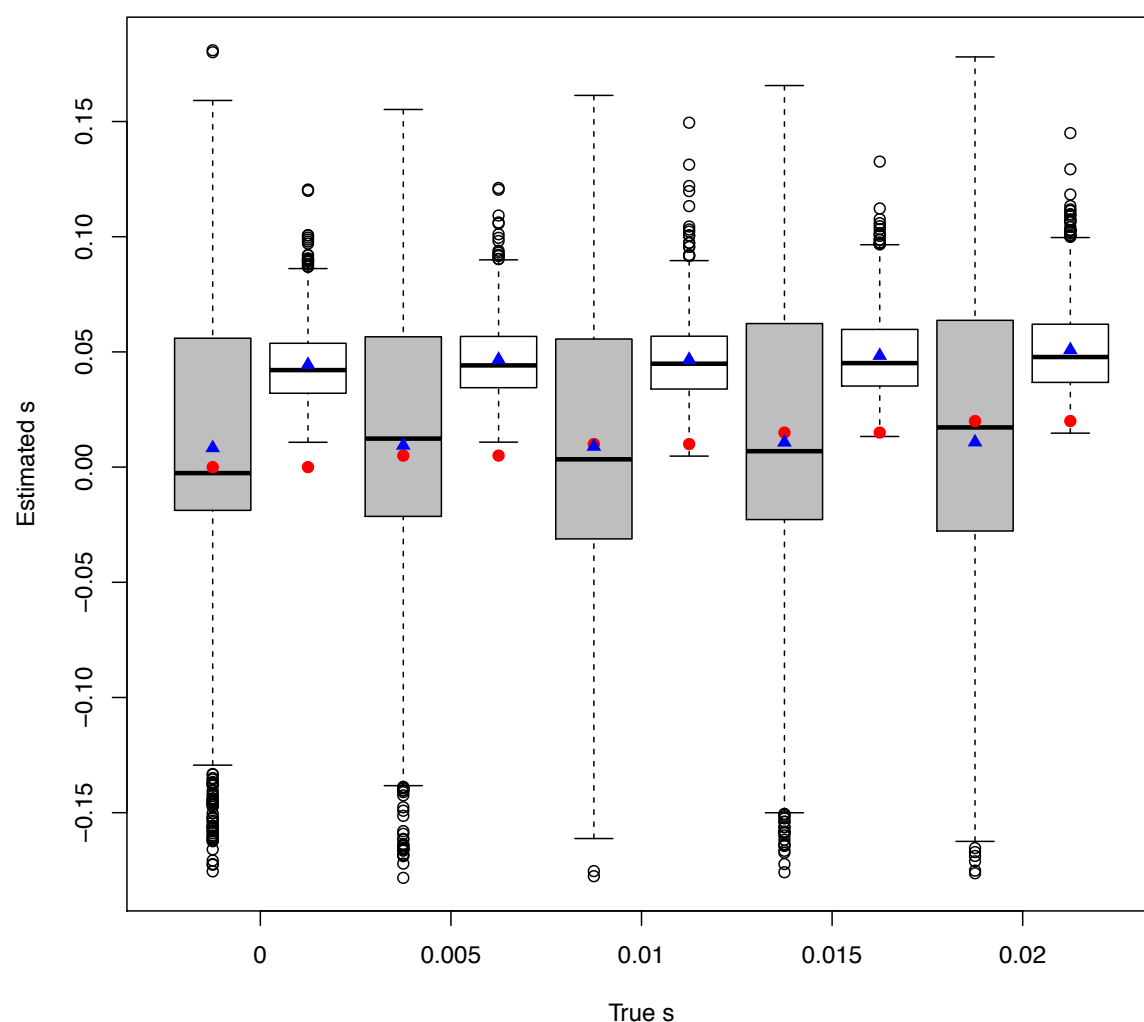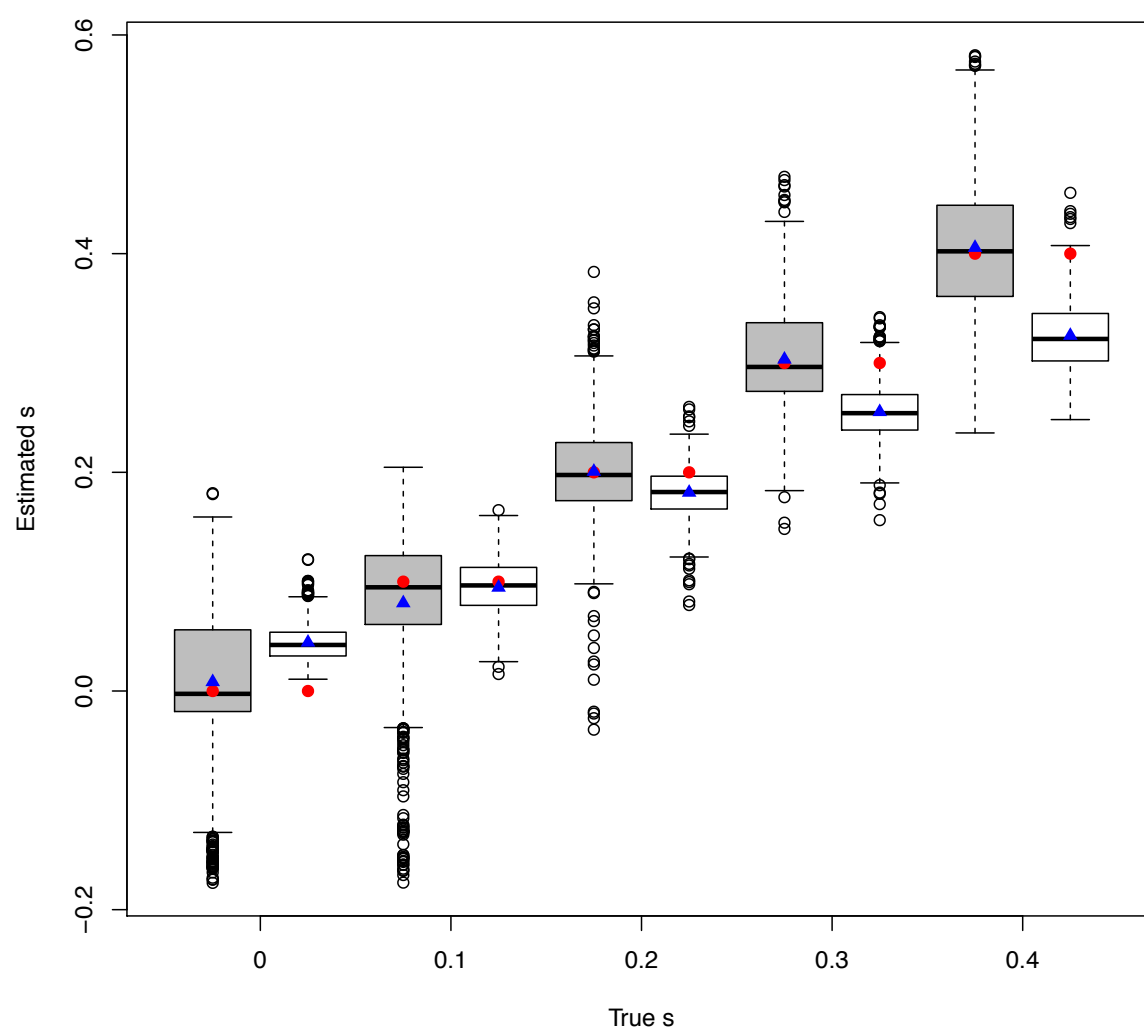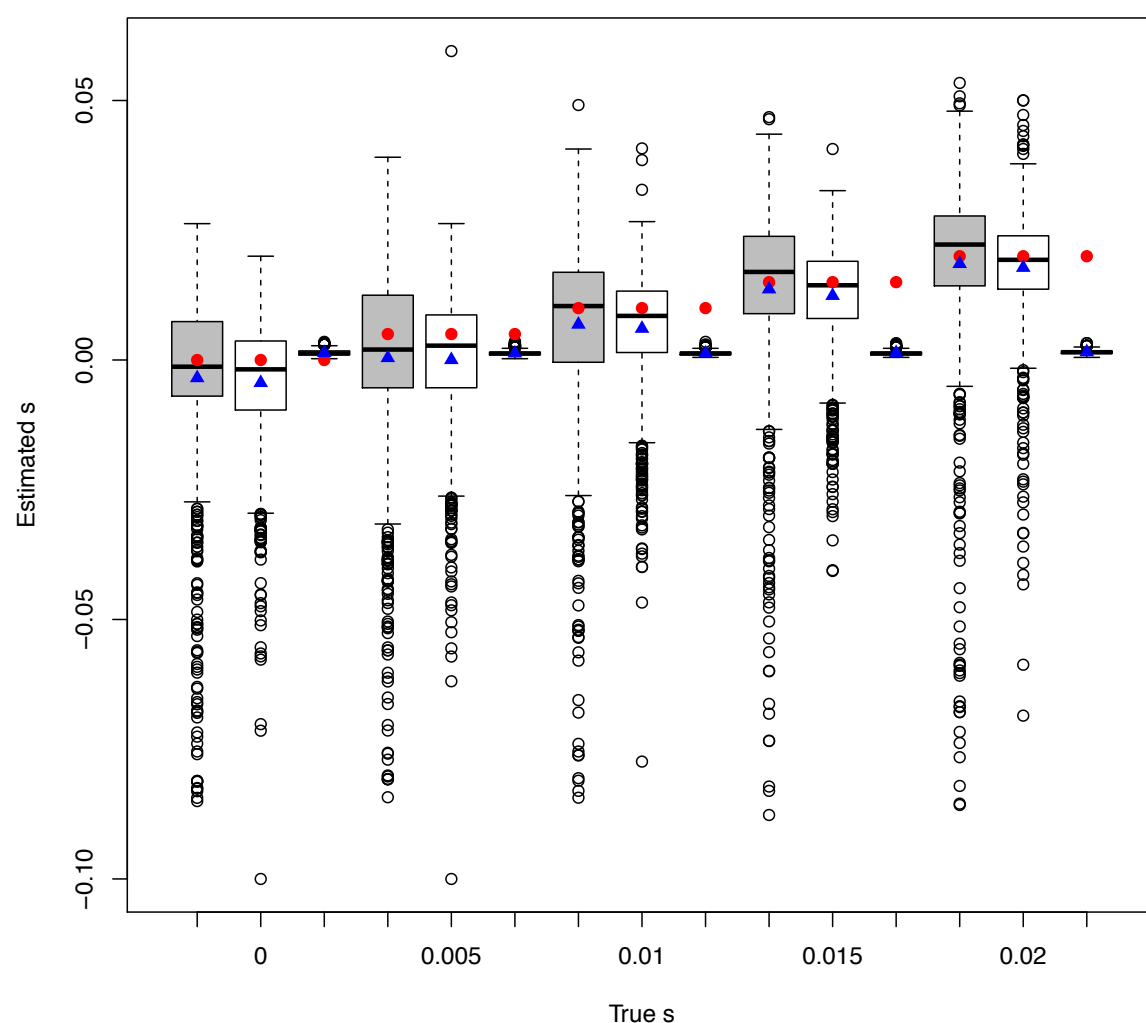
Figure 2. Boxplot for the estimated large selection coefficients from each simulated replicate of the Wright-Fisher diploid model, with N_e=1000, 90 generations and the minimum frequency of 5% at one sampling time point. The gray rectangles correspond to WFABC and the white rectangles to Mathieson and McVean (2013) method. The red circles are the true values for *s*, and the blue triangles are the mean of the estimated *s* values.

Figure 3. Boxplot for the estimated small selection coefficients from each simulated replicate of the Wright-Fisher diploid model, for N_e=1000 and 500 generations. The gray rectangles correspond to WFABC, the white rectangles to the Malaspinas *et al.* (2012) method, and the gold rectangles to the Bollback *et al.* (2008) method. The red circles are the true values of *s*, and the blue triangles are the mean of the estimated *s* values.

Figure 4. Boxplot for the estimated large selection coefficients from each simulation replicate of the Wright-Fisher diploid model, for N_e=1000 and 90 generations. The gray rectangles correspond to WFABC, the white rectangles to the Malaspinas *et al.* (2012) method, and the gold rectangles to the Bollback *et al.* (2008) method. The red circles are the true values for *s*, and the blue triangles are the mean of the estimated *s* values.
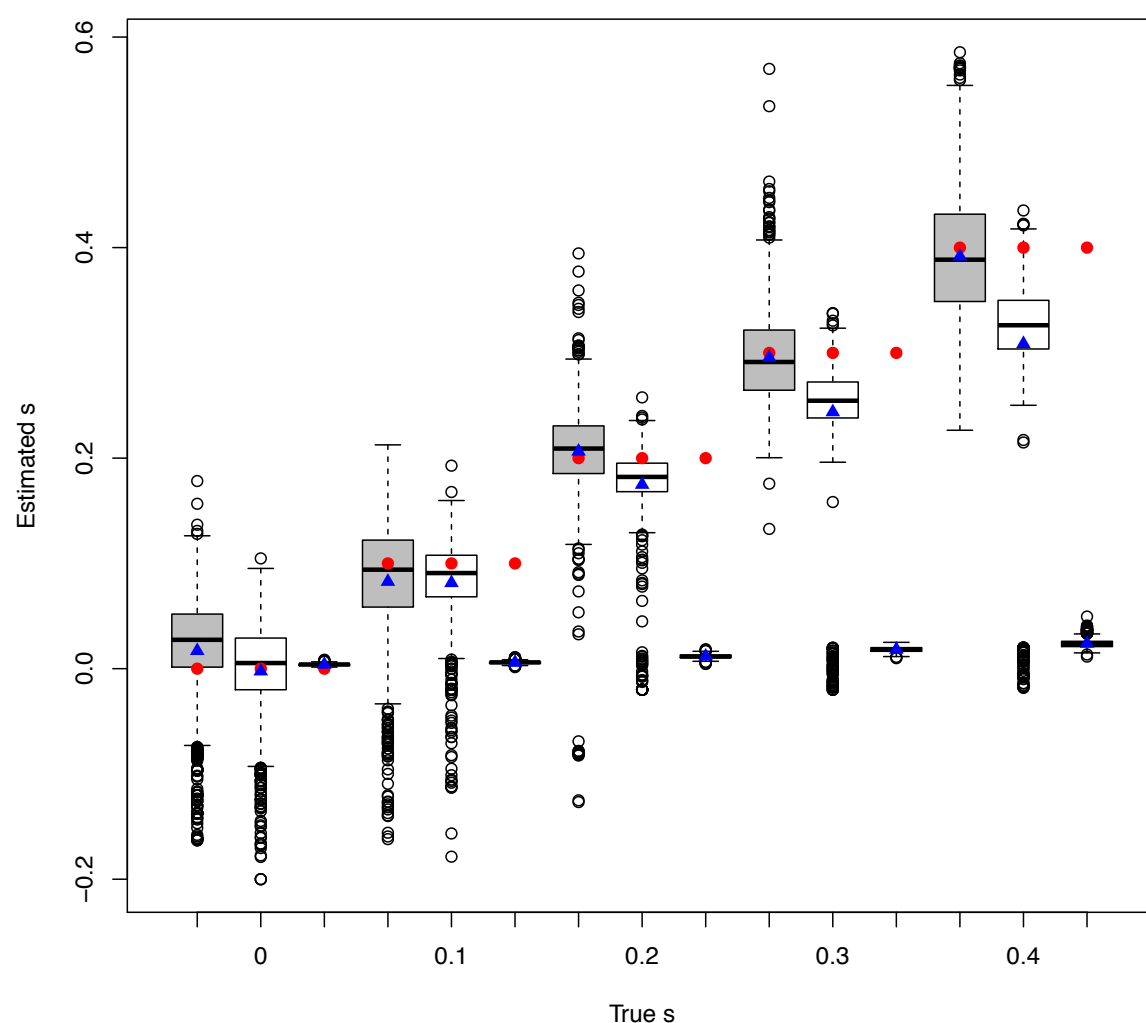
Figure 5. Correlation between the true and the estimated values of the selection coefficients *s* for the multi-locus scenario using WFABC (A) and the Malaspinas *et al.* (2012) method (B). The red line represents the identity (*y=x*) line.
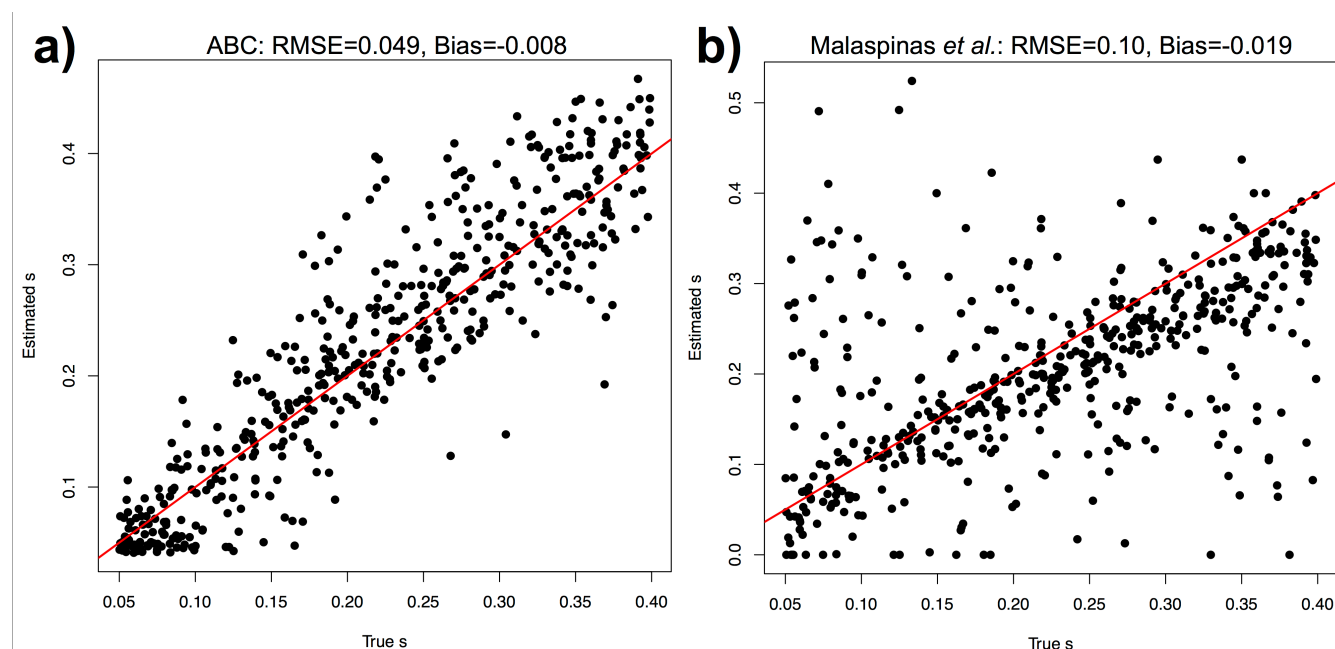
Figure 6. Frequency of the *medionigra* allele in the Cothill *P. dominula* moth population

(Cook and Jones 1996; Jones 2000).
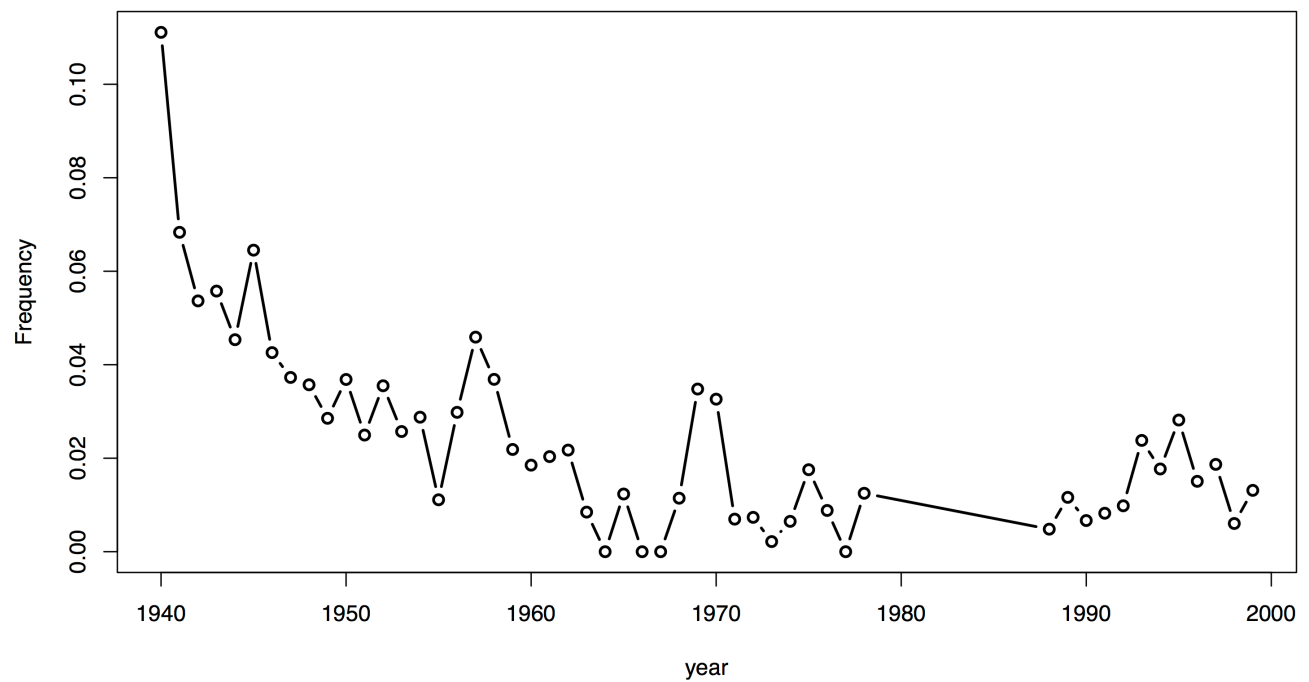
Figure 7. Two-dimensional joint posterior distribution for *s* and *h* for the moth *P. dominula* data using *2Ne=1000* (a) and *2Ne=100* (b). The grey lines delimit two-dimensional α highest posterior density regions for α =0.9 (largest region), 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2 and 0.1 (smallest region).
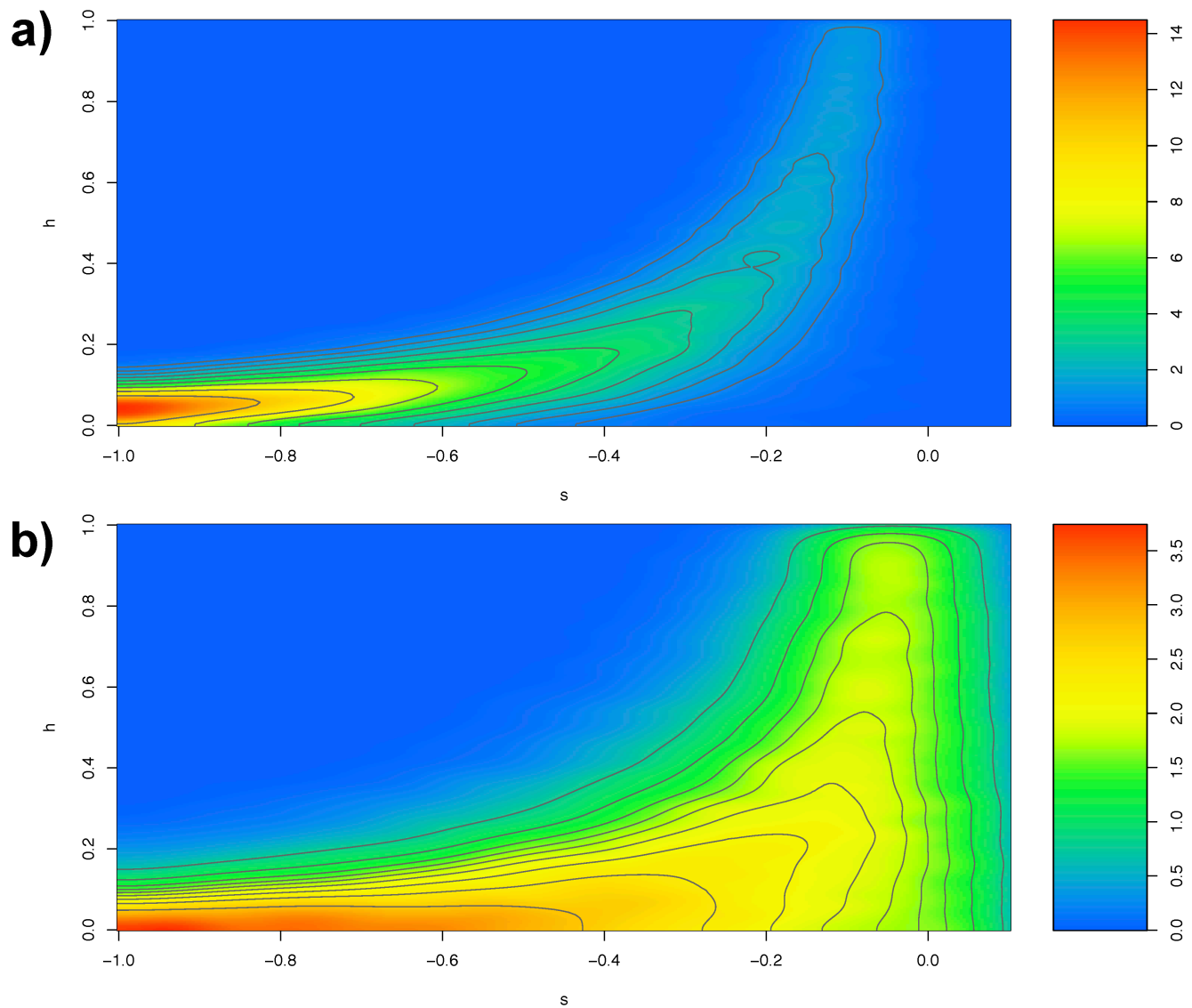
Figure 8. Boxplot for the estimated selection coefficient (*s*, panel a) and dominance ratio (*h*, panel b) from 1000 simulated datasets mimicking the *P. dominula* data. The red circles indicate the true simulated values (*s=-1; h=0.05*), and the blue triangles the mean of the estimated values.
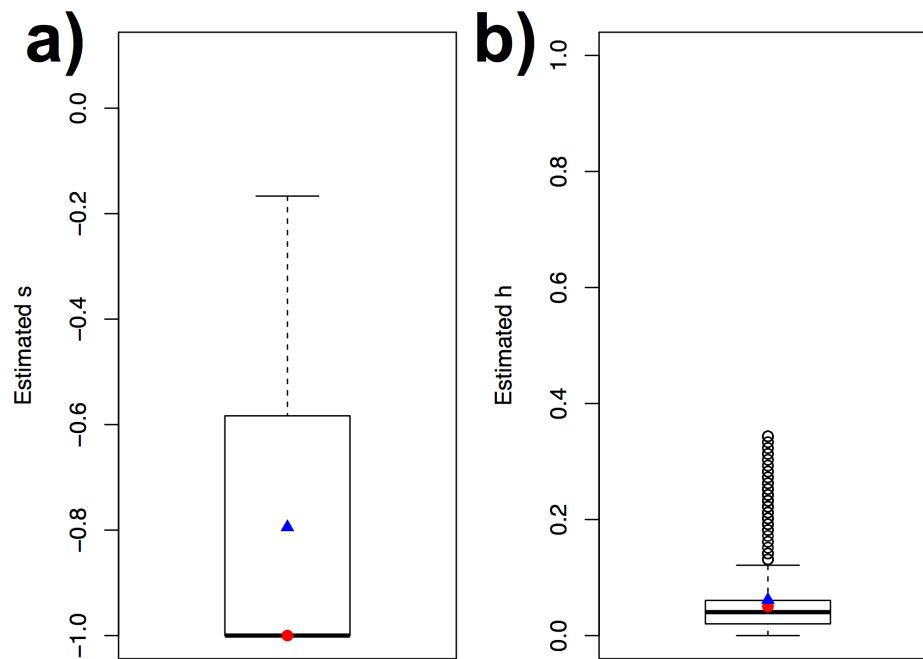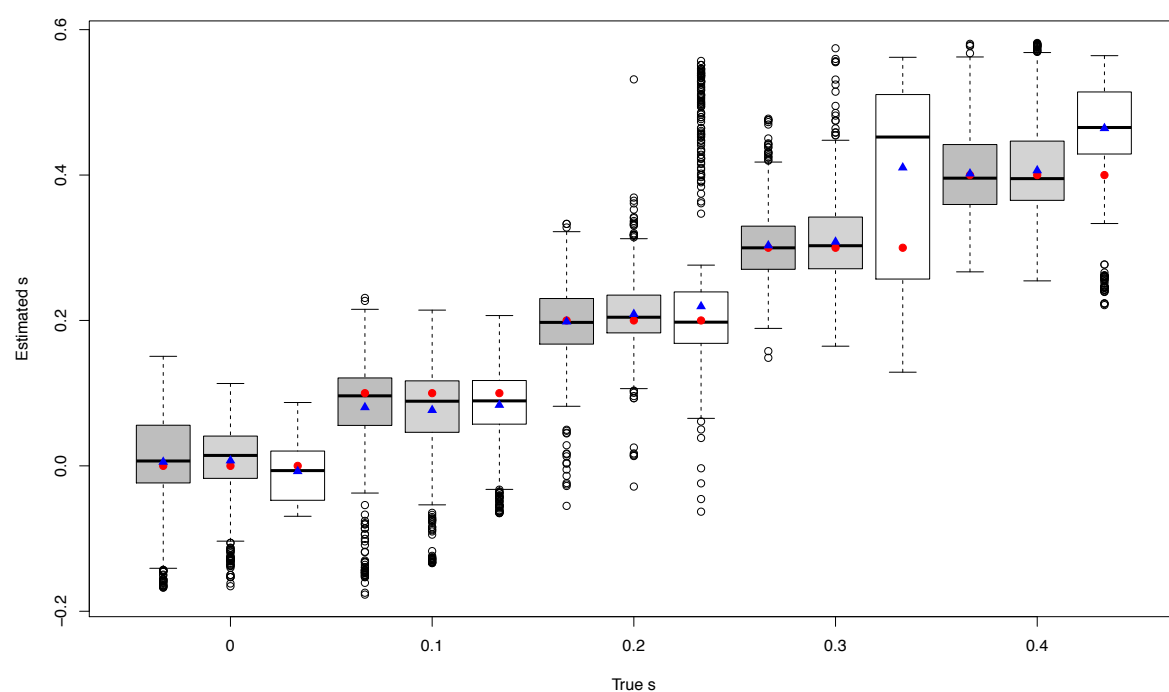
Figure S1. Boxplot for the estimated selection coefficient from WFABC for 3 different sampling time points. Each simulation replicate is from the Wright-Fisher diploid model with N_e=1000 simulated for 90 generations. The dark gray rectangles correspond to 12 sampling time points, the light gray rectangles to 6 sampling time points, and the white rectangles to 2 sampling time points. The red circles are the true values for *s*, and the blue triangles are the mean of the estimated *s* values.

Figure S2. Boxplot for the estimated selection coefficient from WFABC for 3 different sample sizes. Each simulation replicate is from the Wright-Fisher diploid model with N_e=1000 simulated for 90 generations. The dark gray rectangles correspond to 1000 samples, the light gray rectangles to 100 samples, and the white rectangles to 20 samples. The red circles are the true values for *s*, and the blue triangles are the mean of the estimated *s* values.
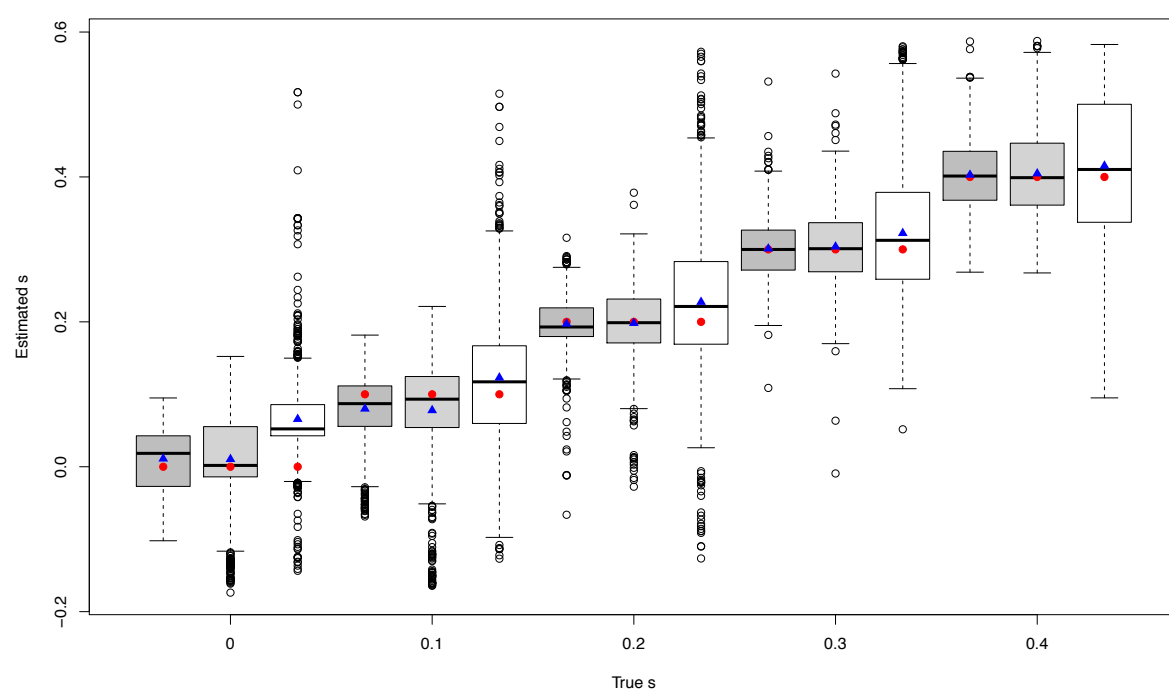
Figure S3. Boxplot for the estimated small *s* from each simulation replicate of the Wright-Fisher diploid model with N_e=1000 simulated for 90 generations and the initial minor allele frequency at 10%. The gray rectangles correspond to WFABC and the white rectangles to the Mathieson *et al.* (2013) method. The red circles are the true values for *s*, and the blue triangles are the mean of the estimated *s* values.
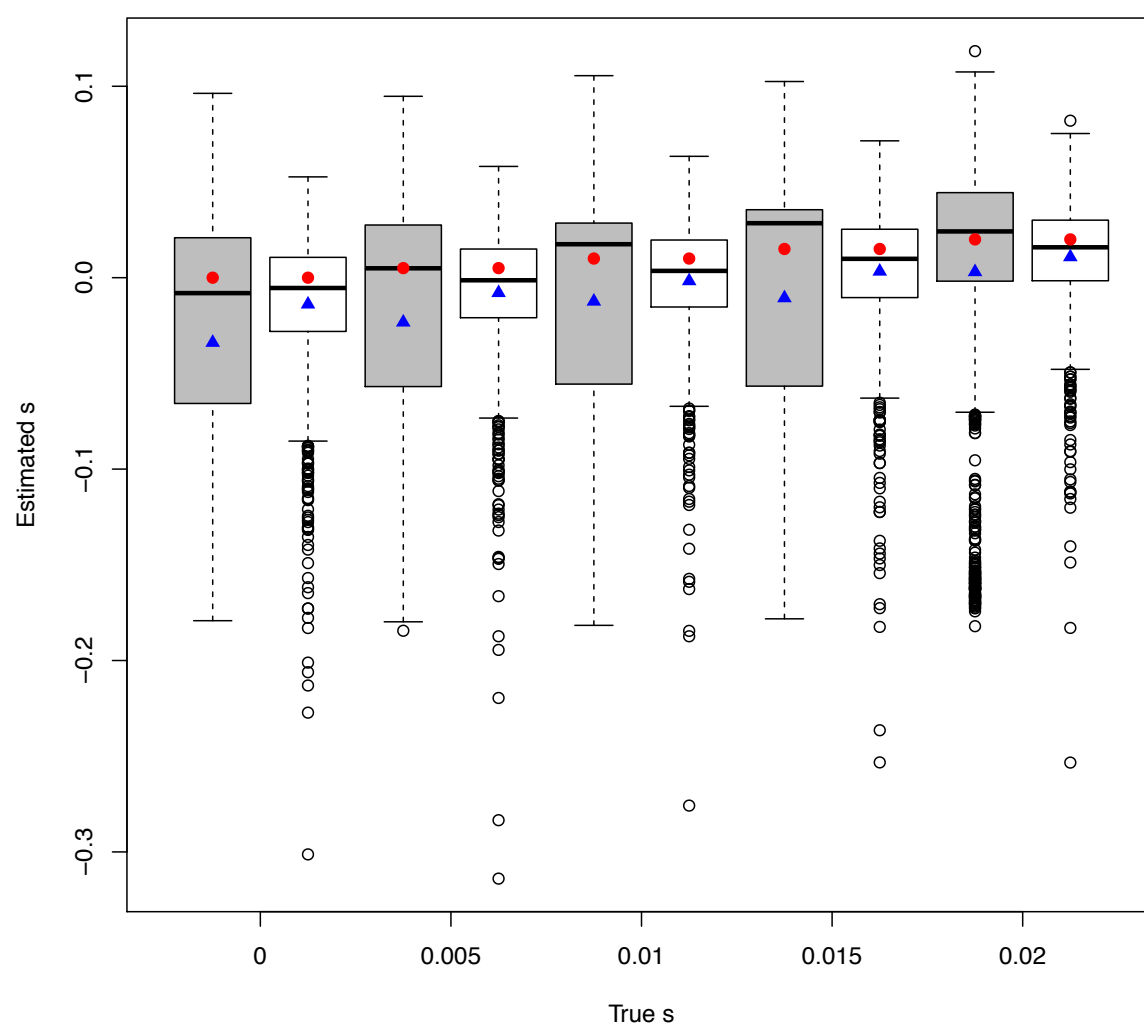
Figure S4. Boxplot for the estimated large *s* from each simulation replicate of the Wright-Fisher diploid model with N_e=1000 simulated for 90 generations and the initial minor allele frequency at 10%. The gray rectangles correspond to WFABC and the white rectangles to the Mathieson *et al.* (2013) method. The red circles are the true values for *s*, and the blue triangles are the mean of the estimated *s* values.
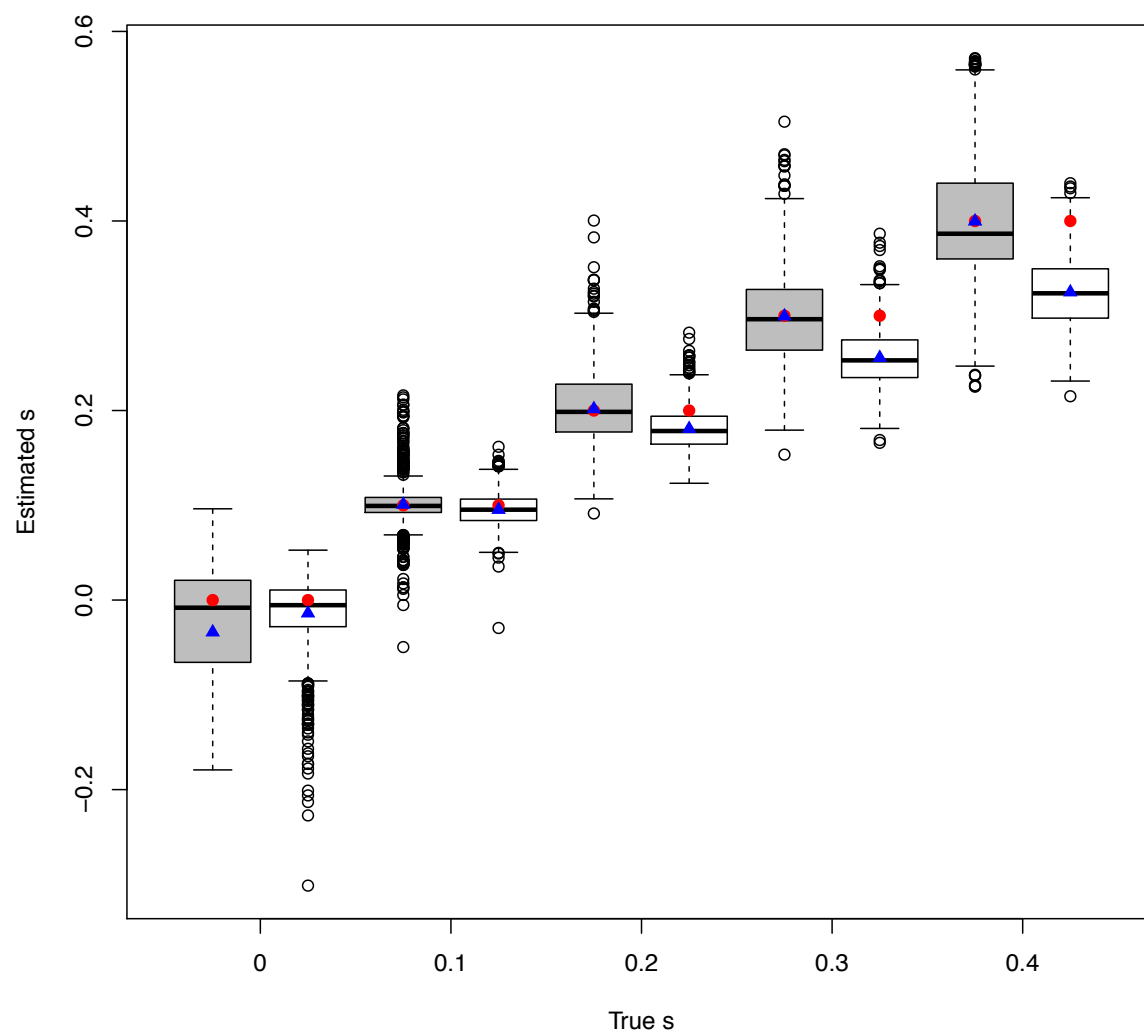
Figure S5. Log likelihood of the estimated selection coefficient for the Bollback *et al.* (2008) method with a small search interval for γ. The input trajectory from the Wright-Fisher diploid model with N_e=1e+07 and *s*=0.4 simulated for 120 generations, and 50 sampled chromosomes.
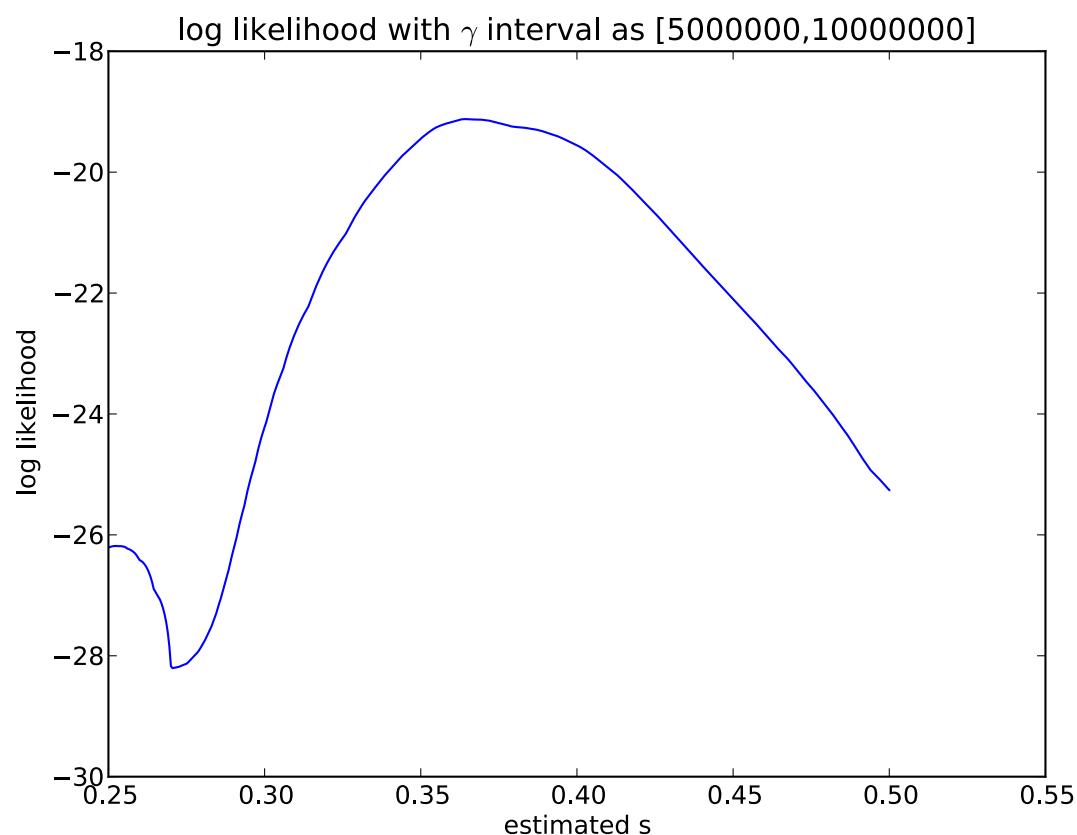
Figure S6. Log likelihood of the estimated selection coefficient for the Bollback *et al.* (2008) method with a large search interval for γ. The input trajectory from the Wright-Fisher diploid model with N_e=1e+07 and *s*=0.4 simulated for 120 generations, and 50 sampled chromosomes.
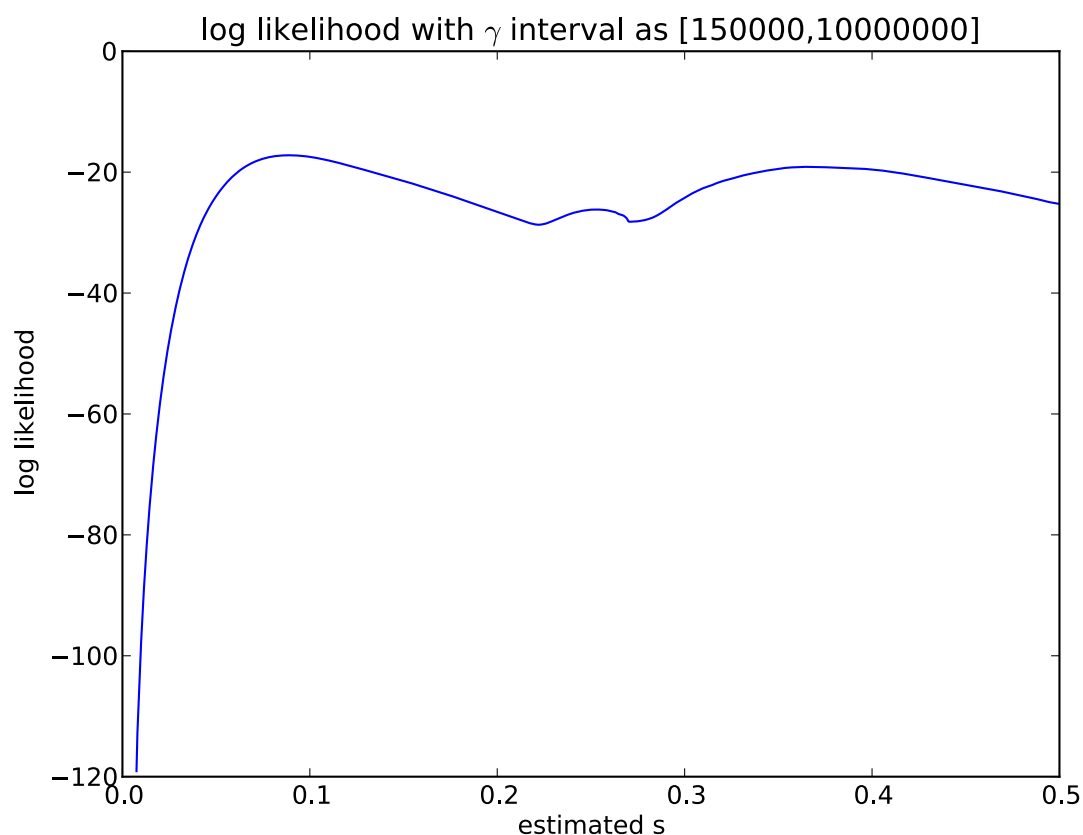
Figure S7. Boxplot for the estimated small selection coefficient from each simulation replicate of the Wright-Fisher diploid model with N_e=200 simulated for 300 generations. The gray rectangles correspond to WFABC, and the white rectangles to the Malaspinas *et al.* (2012) method. The red circles are the true values for *s*, and the blue triangles are the mean of the estimated *s* values.
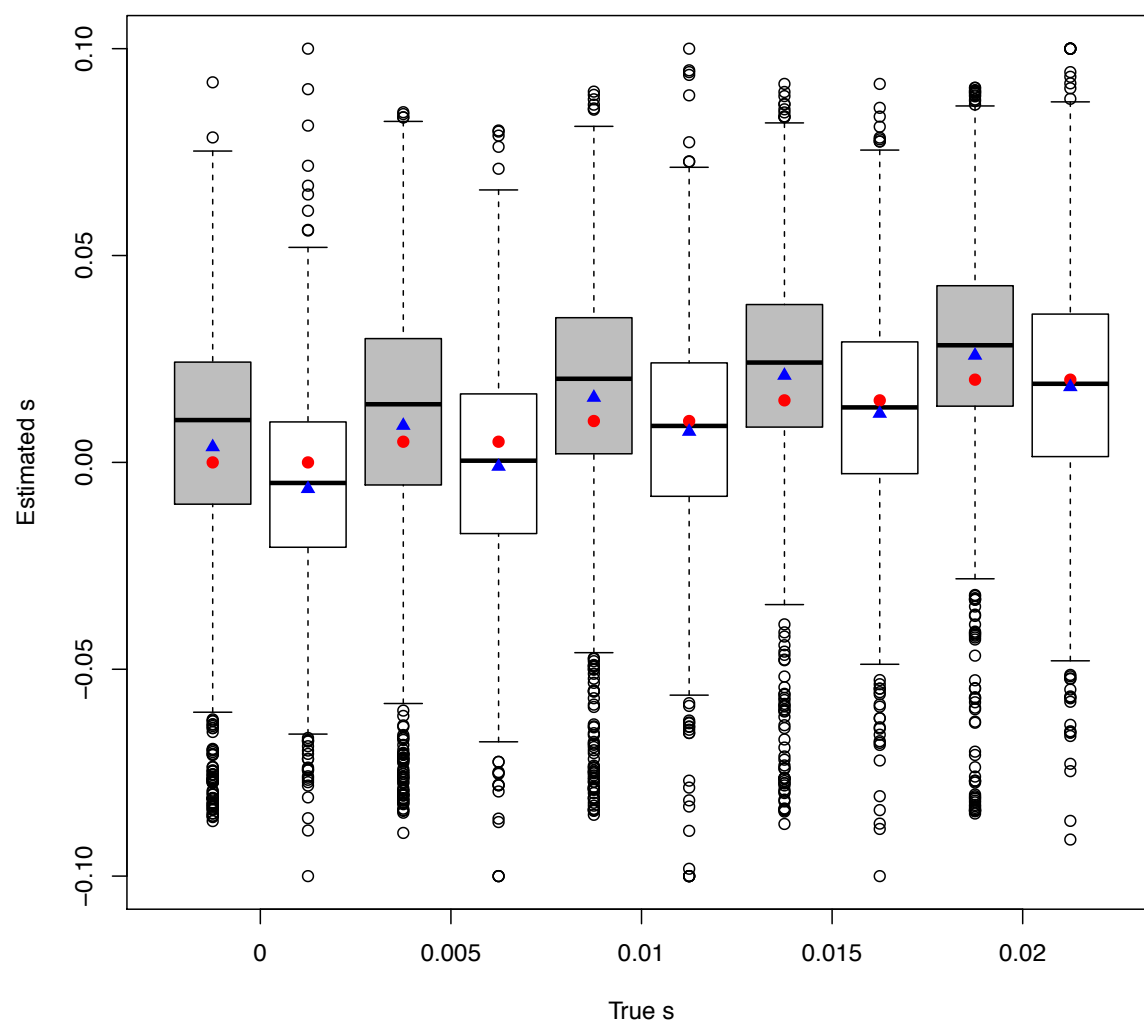
Figure S8. Boxplot for the estimated large selection coefficient from each simulation replicate of the Wright-Fisher diploid model with N_e=200 simulated for 80 generations. The gray rectangles correspond to WFABC, and the white rectangles to the Malaspinas *et al.* (2012) method. The red circles are the true values for *s*, and the blue triangles are the mean of the estimated *s* values.

Figure S9. Boxplot for the estimated small selection coefficient from each simulation replicate of the Wright-Fisher diploid model with N_e=5000 simulated for 500 generations. The gray rectangles correspond to WFABC, and the white rectangles to the Malaspinas *et al.* (2012) method. The red circles are the true values for *s*, and the blue triangles are the mean of the estimated *s* values.
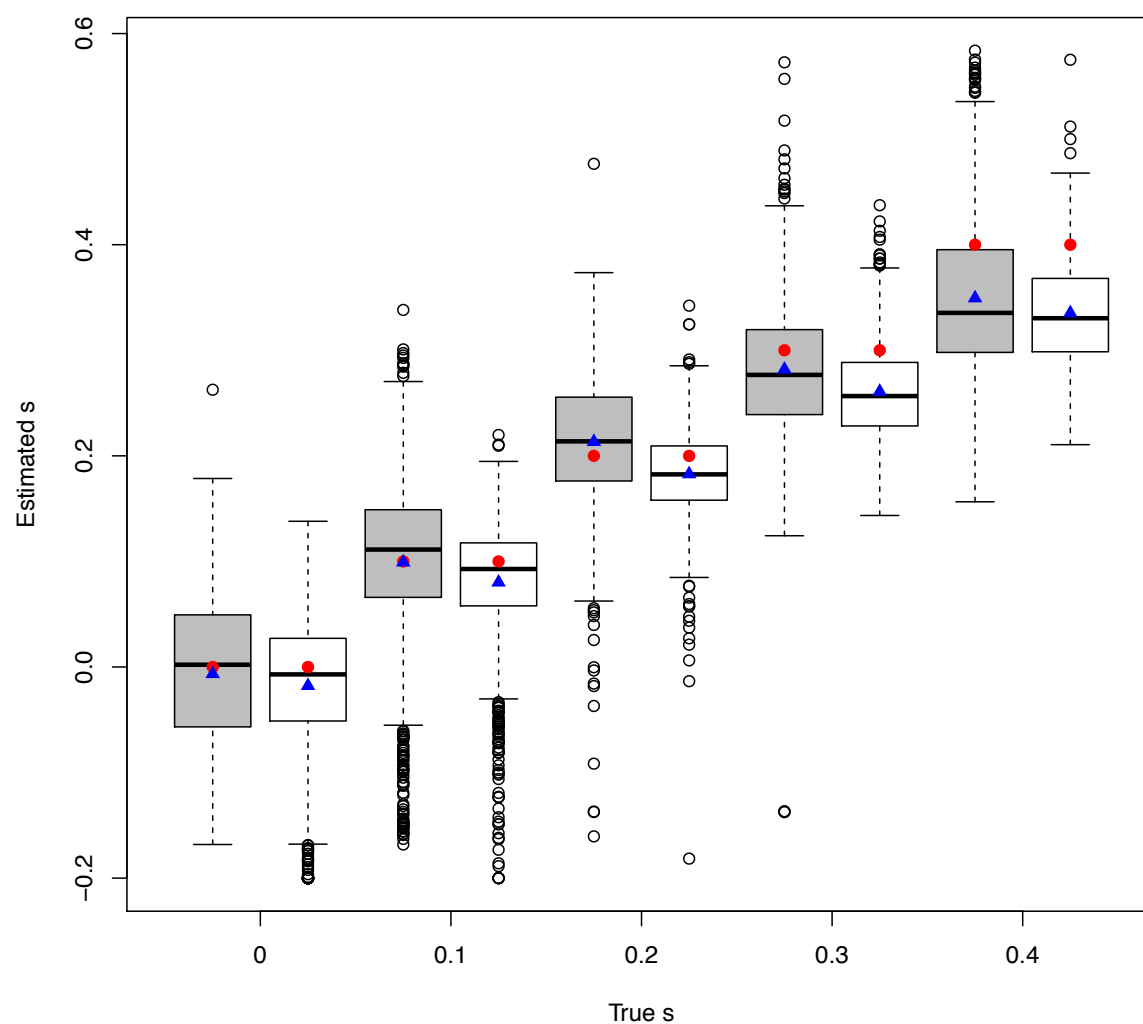
Figure S10. Boxplot for the estimated large selection coefficient from each simulation replicate of the Wright-Fisher diploid model with N_e=5000 simulated for 100 generations. The gray rectangles correspond to WFABC. The red circles are the true values for *s*, and the blue triangles are the mean of the estimated *s* values.

Figure S11. Boxplot for the estimated small selection coefficient from each simulation replicate of the Wright-Fisher haploid model with N_e=1000 simulated for 300 generations. The gray rectangles correspond to WFABC, and the white rectangles to the Malaspinas *et al.* (2012) method. The red circles are the true values for *s*, and the blue triangles are the mean of the estimated *s* values.
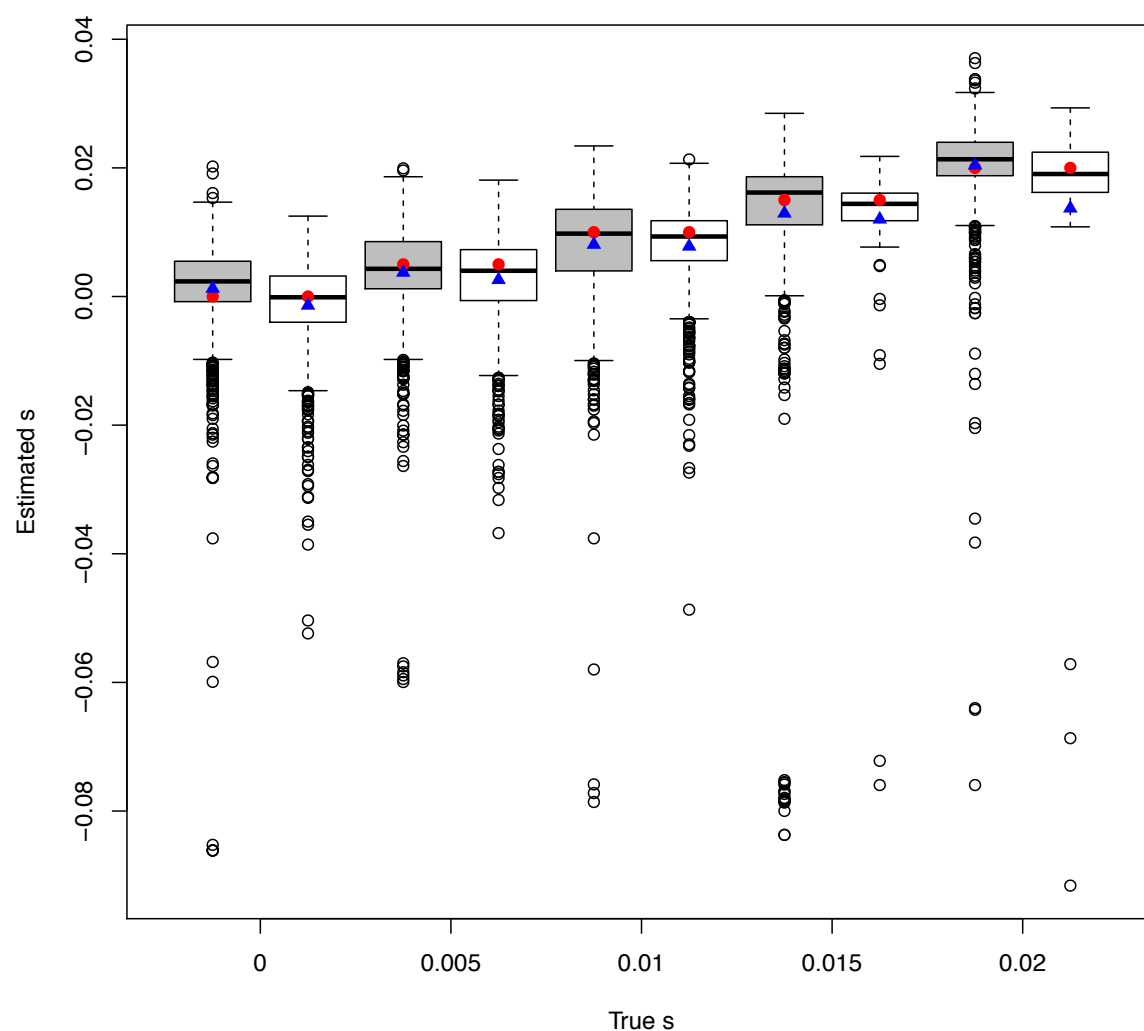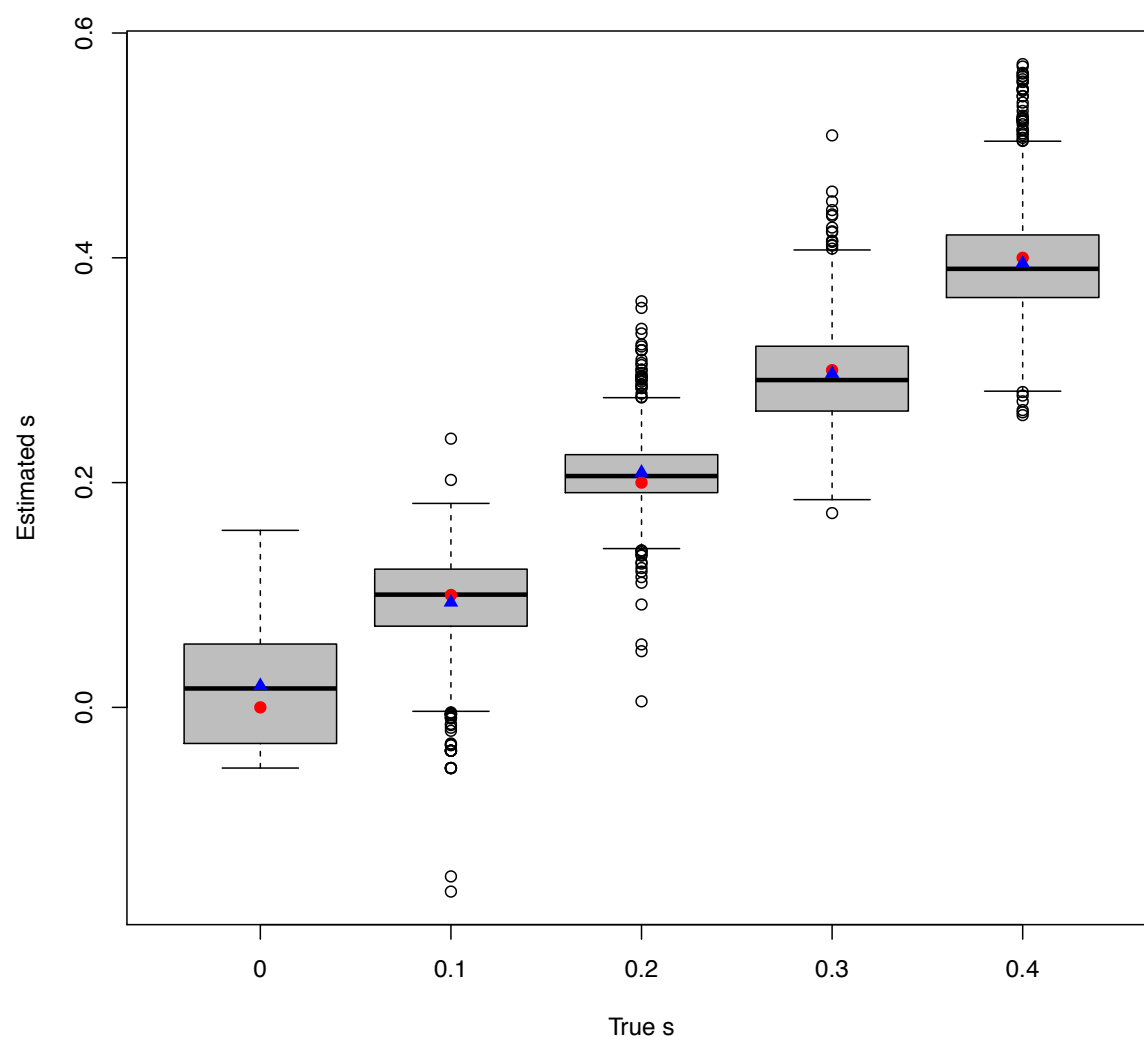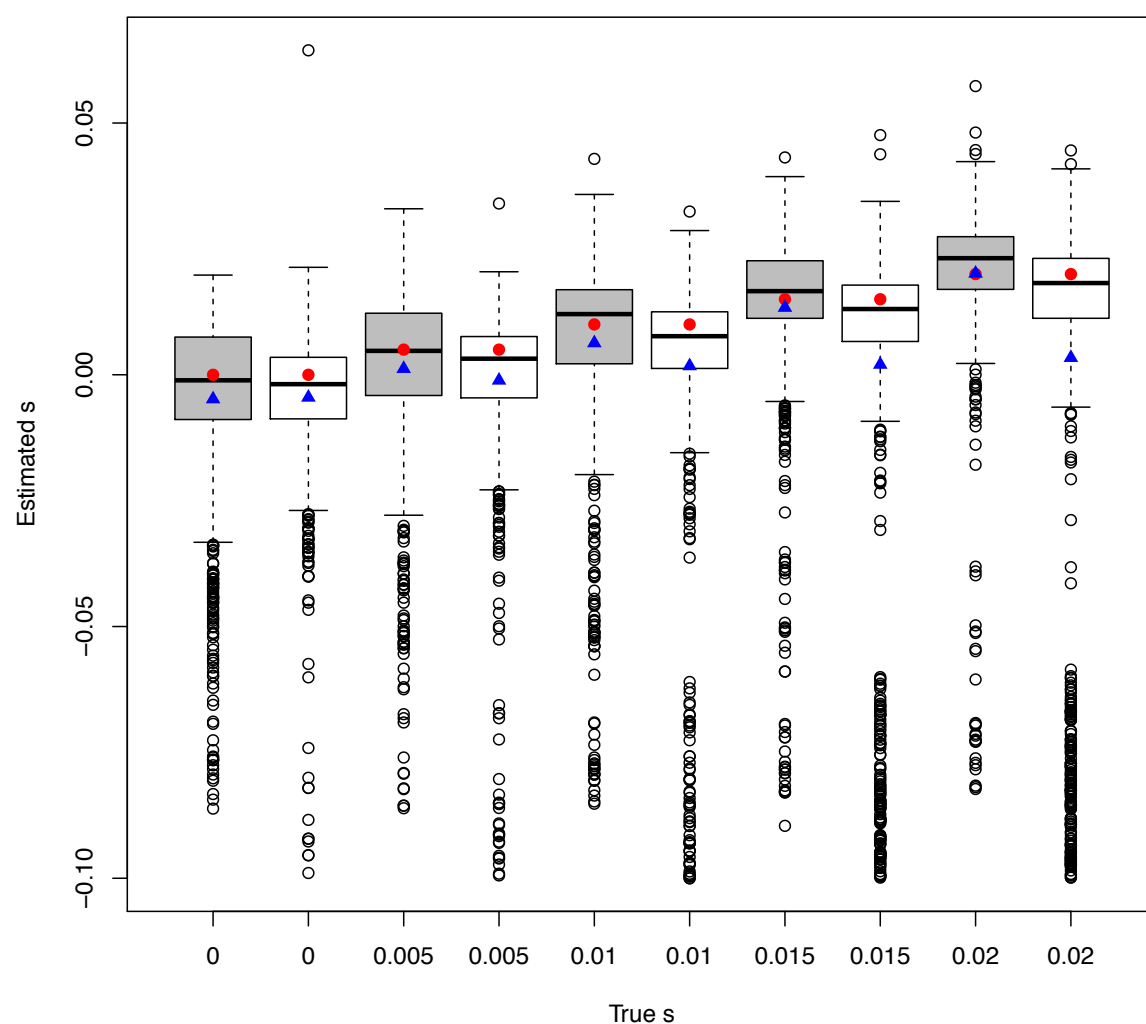
Figure S12. Boxplot for the estimated large selection coefficient from each simulation replicate of the Wright-Fisher haploid model with N_e=1000 simulated for 50 generations. The gray rectangles correspond to WFABC, and the white rectangles to the Malaspinas *et al.* (2012) method. The red circles are the true values for *s*, and the blue triangles are the mean of the estimated *s* values.
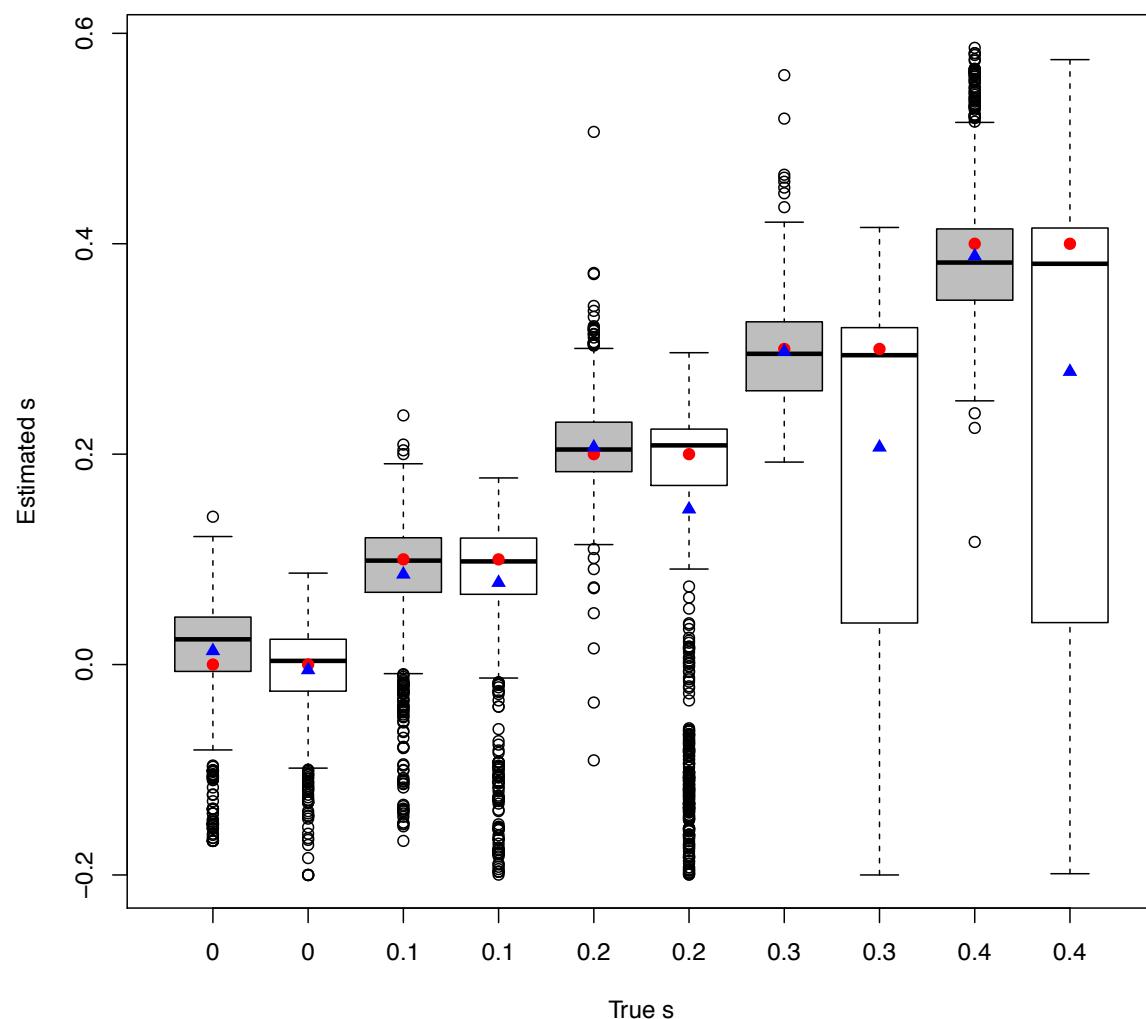
Figure S13. Two-dimensional joint posterior distribution for *s* and *h* for the moth *P. dominula* data using a uniform prior for *2Nₑ* between 100 and 10000. The grey lines delimit two-dimensional α highest posterior density regions for α =0.9 (largest region), 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2 and 0.1 (smallest region).

Table S1. RMSE and bias for the small *s* and N_e=200 scenario for the Wright-Fisher diploid model

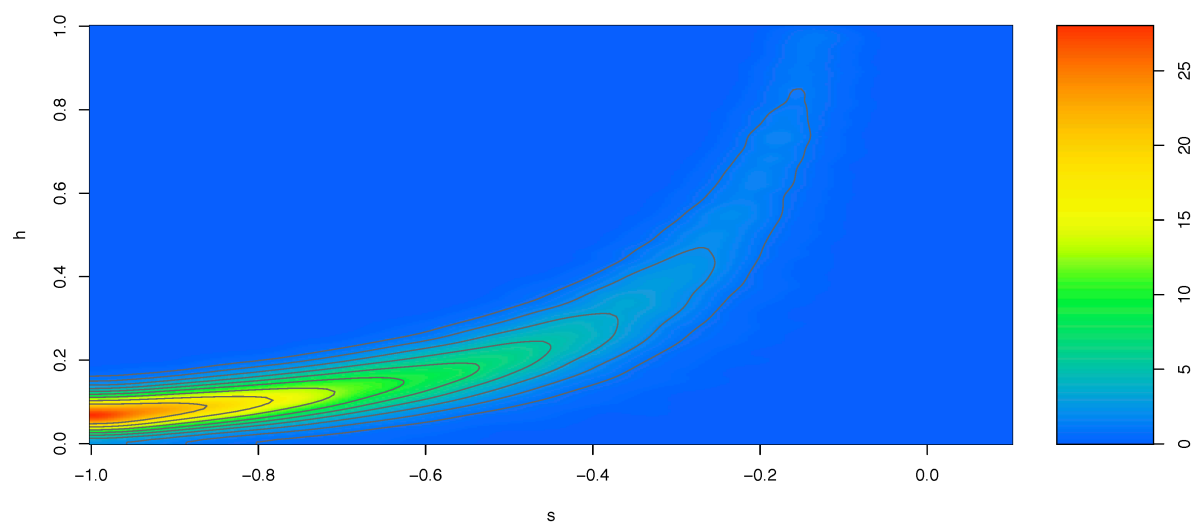|		s=0	s=0.005	s=0.01	s=0.015	s=0.02
Bias	WFABC	0.0037	0.0039	0.0057	0.0060	0.0058
	Malaspinas *et al.* (2012)	-0.0064	-0.0060	-0.0026	-0.0032	-0.0018
RMSE	WFABC	0.031	0.032	0.032	0.030	0.031
	Malaspinas *et al.* (2012)	0.026	0.027	0.028	0.027	0.028

Table S2. RMSE and bias for the big *s* and N_e=200 scenario for the Wright-Fisher diploid model

		s=0	s=0.1	s=0.2	s=0.3	s=0.4
Bias	WFABC	-0.0065	-0.00081	0.013	-0.018	-0.051
	Malaspinas *et al.* (2012)	-0.018	-0.020	-0.017	-0.039	-0.065
RMSE	WFABC	0.075	0.081	0.066	0.067	0.087
	Malaspinas *et al.* (2012)	0.068	0.065	0.045	0.060	0.083

Table S3. RMSE and bias for the small *s* and N_e=5000 scenario for the Wright-Fisher diploid model

		s=0	s=0.005	s=0.01	s=0.015	s=0.02
Bias	WFABC	0.0012	-0.0013	-0.0019	-0.0021	0.00040
	Malaspinas *et al.* (2012)	-0.0014	-0.0024	-0.0022	-	-
RMSE	WFABC	0.0094	0.0082	0.0091	0.014	0.0078
	Malaspinas *et al.* (2012)	0.0073	0.0073	0.0069	-	-

Table S4. RMSE and bias for the big *s* and N_e=5000 scenario for the Wright-Fisher diploid model

		s=0	s=0.1	s=0.2	s=0.3	s=0.4
Bias	WFABC	0.019	-0.0066	0.0090	-0.0031	-0.0049
	Malaspinas *et al.* (2012)	-	-	-	-	-
RMSE	WFABC	0.052	0.044	0.034	0.045	0.053
	Malaspinas *et al.* (2012)	-	-	-	-	-

Table S5. RMSE and bias for the small *s* and N_e=1000 scenario for the Wright-Fisher haploid model

		s=0	s=0.005	s=0.01	s=0.015	s=0.02
Bias	WFABC	-0.0048	-0.0038	-0.0037	-0.0017	0.00012
	Malaspinas *et al.* (2012)	-0.0045	0.0061	-0.0083	-0.013	-0.017
RMSE	WFABC	0.019	0.017	0.020	0.018	0.016
	Malaspinas *et al.* (2012)	0.015	0.018	0.025	0.035	0.041

Table S6. RMSE and bias for the big *s* and N_e=1000 scenario for the Wright-Fisher haploid model

		s=0	s=0.1	s=0.2	s=0.3	s=0.4
Bias	WFABC	0.013	-0.014	0.0065	-0.00247	-0.012
	Malaspinas *et al.* (2012)	-0.0052	-0.022	-0.052	-0.094	-0.12
RMSE	WFABC	0.055	0.059	0.041	0.049	0.064
	Malaspinas *et al.* (2012)	0.046	0.076	0.14	0.20	0.25