Prediction and estimation of effective population size

Jinliang Wang[1], Enrique Santiago[2], Armando Caballero[3]

[1] *Institute of Zoology, Zoological Society of London, London NW1 4RY, United Kingdom*

[2] *Departamento de Biología Funcional, Facultad de Biología, Universidad de Oviedo, 33071 Oviedo, Spain*

[3] *Departamento de Bioquímica, Genética e Inmunología, Facultad de Biología, Universidad de Vigo, 36310 Vigo, Spain*

16   *Corresponding author:*

17   Jinliang Wang

18   Institute of Zoology

19   Regent's Park

20   London NW1 4RY

21   United Kingdom

22   Tel: 0044 20 74496620

23   Fax: 0044 20 75862870

24   Email: jinliang.wang@ioz.ac.uk

25

ABSTRACT

26

27 Effective population size (N_e) is a key parameter in population genetics. It has important

28 applications in evolutionary biology, conservation genetics, and plant and animal breeding, because

29 it measures the rates of genetic drift and inbreeding and affects the efficacy of systematic

30 evolutionary forces such as mutation, selection and migration. We review the developments in

31 predictive equations and estimation methodologies of effective size. In the prediction part, we focus

32 on the equations for populations with different modes of reproduction, for populations under

33 selection for unlinked or linked loci, and for the specific applications to conservation genetics. In

34 the estimation part, we focus on methods developed for estimating the current or recent effective

35 size from molecular marker or sequence data. We discuss some underdeveloped areas in predicting

36 and estimating N_e for future research.

37

38

Introduction

40

41 The concept of effective population size, introduced by Sewall Wright (1931, 1933), is central to

42 plant and animal breeding (Falconer and Mackay, 1996), conservation genetics (Frankham *et al*.,

43 2010; Allendorf *et al*., 2013), and molecular variation and evolution (Charlesworth and

44 Charlesworth, 2010), as it quantifies the magnitude of genetic drift and inbreeding in real-world

45 populations. A substantial number of extensions to the basic theory and predictions were made

46 since the seminal work of Wright, with main early developments by James Crow and Motoo

47 Kimura (Kimura and Crow, 1963a; Crow and Kimura, 1970) and later by a list of contributors.

48 Several review papers (Crow and Denniston, 1988; Caballero, 1994; Wang and Caballero, 1999;

49 Nomura, 2005a), and population genetics books (Fisher, 1965; Wright, 1969; Ewens, 1979;

50 Nagylaki, 1992) have summarised the existing theory in predicting the effective size of a population

51 at different spatial and time scales under various inheritance modes and demographies.

52 Comparatively, methodological developments (reviewed by Schwartz *et al*., 1999; Beaumont,

53 2003a; Wang, 2005; Palstra and Ruzzante, 2008; Luikart *et al*., 2010; Gilbert and Whitlock, 2015)

54 in estimating the effective size of natural populations from genetic data lag behind, but are

55 accelerating in the past decade thanks to the rapid developments of molecular biology.

56   The classical developments of effective population size theory are based on the rate of

57 change in gene frequency variance (genetic drift) or the rate of inbreeding. The effective population

58 size is defined in reference to the Wright-Fisher idealised population, i.e. a hypothetical population

59 with very simplifying characteristics where genetic drift is the only factor in operation, and the

60 dynamics of allelic and genotypic frequencies across generations merely depend on the population

61 census (*N*) size. The effective size of a real population is then defined as the size of an idealised

62 population which would give rise to the rate of inbreeding and the rate of change in variance of

63 gene frequencies actually observed in the population under consideration, which correspond to the

64 so-called *inbreeding and variance effective sizes*, respectively (Crow and Kimura, 1970).

65   Predictions of the effective population size can also be obtained from the largest nonunit

66 eigenvalue of the transition matrix of a Markov Chain which describes the dynamics of allele

67 frequencies. Such derived effective size is called *eigenvalue effective size* (see Ewens, 1979, p. 104-

68 112), which is equivalent to the *random extinction effective size* (Crow, 1954; see also Haldane,

69 1939). The transition matrix can be written for many genetic models, and is particularly useful for

70 complex scenarios such as populations varying in size, having age structures or being subject to

71 demographic changes (e.g., Wang and Pollak, 2002; Charlesworth, 2001; Pollak, 2002; Engen *et*

72 *al*., 2005). A less often used approach is that for the *mutation effective size*, defined by the

73 probability of identity in state of genes rather than identity by descent under an infinite allele model

74 of mutations with a defined mutation rate (Whitlock and Barton, 1997).

75   Later developments based on coalescence theory (Wakeley, 2008) have also proved to be

76 useful in the prediction of effective population size, particularly in the evolutionary context for

77 predicting genetic variability at the molecular level (Charlesworth, 2009; Nicolaisen and Desai,

78 2012, 2013). The coalescence theory states that the chance of coalescence of any two random gene

79 copies in one generation time is $1/2N$, which is the same as the rate of increase in identity by

80 descent occurred from one generation to the next one. Thus, the probability of coalescence *t*

81 generations ago is $[1 - (1/2N)]^{t-1}(1/2N)$. Therefore, the average time of coalescence of two

82 randomly chosen genes is $T = \left(1/2N\right) \sum_{t=1}^{\infty} t \left[1 - \left(1/2N\right)\right]^{t-1} = 2N$. The *coalescent effective*

83 *population size* refers to the expected time of coalescence *T*, in generations, of gene copies such that

84 $T = 2N_e$ (Nordborg and Krone, 2002; Wakeley and Sargsyan, 2009).

85   In this paper, we present a general overview of the main developments for predicting the

86 effective population size (N_e). The review does not attempt to be exhaustive, and some of the

87 material mentioned in previous reviews will not be repeated. We mainly focus on populations with

88 different modes of reproduction, populations under selection, and populations under genetic

89 management in captive breeding conservation programmes, complementing previous reviews and

90 adding material not covered or only partially covered by them. We also review the developments in

91 estimating contemporary effective sizes from genetic marker data, focusing on the estimation

92 principles and ignoring the technical details that were covered in the original papers. The

93 underlying assumptions, application scopes, robustness and accuracies of different estimation

94 methods are discussed and compared.

95

Prediction of the effective population size

96

97

98 In this section, we will summarize the main predictive equations for the asymptotic effective

99 population size reached after a number of generations in a regular breeding system. In this case, all

100 of the above approaches generally lead to the same predictive equations of N_e except for a few

101 particular scenarios. For example, in a regular breeding system for an undivided population the

102 asymptotic inbreeding and variance effective sizes converge. Only in situations such as when the

103 population is subdivided permanently in independent sublines with completely independent

104 pedigrees (Wang, 1997a, b), or when the population is decreasing or increasing in size, these types

105 of N_e will differ permanently. In fact, many of the equations shown below have been derived by two

106 or more of the above approaches, and we will only mention some of these. For clarity and a better

107 understanding of the main principles, several simplifying assumptions will also be made in this

108 prediction section. Unless otherwise stated, we will assume that populations do not change size

109 through time and are large enough so that second order terms of $1/N_e$ can be safely ignored. These

110 terms are generally of little relevance, but make the derivations and the N_e equations rather

111 cumbersome. Finally, a single undivided population with discrete generations under a regular

112 breeding scheme will be generally assumed unless otherwise indicated, so that prediction equations

113 refer to asymptotic (Caballero, 1994) effective population sizes.

114

115 *Populations with different modes of reproduction*

116

117 As a starting point, we consider the simple equation derived by Wright (1938) which takes account

118 of the variance of the contributions from parents to progeny (S_k^2) in a population of constant size N,

119 $N_e = \frac{4N}{2+S_k^2}$.   (1)

120 This expression also assumes a population either containing only hermaphrodites or comprising

121 equal numbers of males and females, diploid autosomal inheritance, and random mating (including

122 selfing for hermaphrodites). In eqn (1), the term S_k^2 accounts for the genetic drift caused by the

123 variable contributions among parents, whereas the first term "2" in the denominator accounts for the

124 genetic drift caused by the Mendelian segregation of heterozygotes (i.e. the drift in allele frequency

125 arising from the fact that the progeny from a heterozygote can alternatively receive one or the other

126 allele). It can also be seen as the variance in contributions between paternal and maternal genes at a

127 locus within an individual or part of the variance in contribution between grandparents (the term δ^2

128 in eqn 2 of Wang and Hill [2000]).

129   An illustrative generalisation of eqn (1) to the case of different numbers of males (N_m) and

130   females (N_f) is

131    $$N_e = \frac{16 N_m N_f \,/\,(N_m + N_f)}{2 + S_k^2}\,, \tag{2}$$

132   with $S_k^2 = \dfrac{N_f}{N_m + N_f}\left[S_{mm}^2 + 2\left(\dfrac{N_m}{N_f}\right) S_{mm,mf} + \left(\dfrac{N_m}{N_f}\right)^2 S_{mf}^2 \right]$

133    $$+ \frac{N_m}{N_m + N_f}\left[S_{ff}^2 + 2\left(\frac{N_f}{N_m}\right) S_{fm,ff} + \left(\frac{N_f}{N_m}\right)^2 S_{fm}^2 \right], \tag{3}$$

134   where S_{xy}^2 is the variance of the number of offspring of sex y from parents of sex x, and $S_{xm,xf}$ is

135   the covariance of the numbers of male and female offspring from parents of sex x. Eqn (2) is the

136   same as that derived by Hill (1979), although it is expressed in a different form. It reduces to the

137   classical equation of Wright (1933, 1939) for a Poisson distribution of progeny number (i.e.

138    $S_{xy}^2 = N_y / N_x, S_{xm,xf} = 0$ for sex $x, y = m, f$),

139    $$N_e = \frac{4 N_m N_f}{N_m + N_f}\,, \tag{4}$$

140   which shows that unequal numbers of males and females in a population introduce a systematic

141   variance in contribution between male and female parents and thus a reduction in effective size.

142   Predictive formulae for the effective size of X-linked genes were originally given by Wright

143   (1933) and later extended by other authors (see Caballero, 1995). Later developments have also

144   been made for Y-linked and maternally transmitted genes (Charlesworth, 2001; Laporte and

145   Charlesworth, 2002; Evans and Charlesworth, 2013).

146   The generalisation of eqn (1) to the case of a partially selfed population (in which there is

147   partial selfing with proportion β, random mating otherwise), is

148    $$N_e = \frac{4N}{2(1-\alpha) + S_k^2(1+\alpha)}\,, \tag{5}$$

149   (Crow and Morton, 1955), where

150    $$\alpha = \beta \,/\, (2 - \beta) \tag{6}$$

151   (Haldane, 1924) quantifies the deviation from Hardy-Weinberg equilibrium or the correlation of

152   genes within individuals relative to the genes taken at random from the population (Wright, 1969).

153   The value of α in a large random mating population is approximately zero (slightly negative when

154   second order terms are considered; see eqn (23) below and Wang, 1996a). For the case of biparental

155   inbreeding such as partial full-sib mating in dioecious species, the expression is the same as (5)

156   except that the term $(1 + \alpha)$ should be replaced by $(1 + 3\alpha)$ (Caballero and Hill, 1992). The

157   equilibrium value of α for biparental inbreeding is also different from (6) (e.g. Ghai, 1969).

158   The generalisation of eqn (5) for different numbers of males and females was made by

159   Wang (1996b), adding some covariance terms not considered by Caballero (1994; eqn 17). Wang's

160   (1996b) equation also allows for different numbers of males and females varying over generations.

161   An excess of inbred matings ($\alpha > 0$ in eqn [5]) has the effect of increasing the correlation of

162   genes within individuals and decreasing the frequency of heterozygotes by a factor of α. It results in

163   a decrease in the genetic drift due to a decrease in Mendelian segregation variance, and an increase

164   in the genetic drift due to an increase in the variance of contributions among individuals. Compared

165   with random mating ($\alpha = 0$), whether an excess ($\alpha > 0$) or a deficit ($\alpha < 0$) of inbred matings may

166   increase or decrease N_e depends on the variance in family size, S_k^2. For the case of partial selfing

167   (eqn 5), inbreeding ($\alpha > 0$) increases N_e when $S_k^2 < 2$ and decreases N_e when $S_k^2 > 2$. At exactly

168    $S_k^2 = 2$, selfing has no effect on N_e.

169   Predictions of the effective size for X-linked genes in non-random mating populations were

170   given by Wang (1996c). Nomura (2002a, 2005b) also provided predictions of the effective size for

171   a variety of mating systems in animals (see also Balloux and Lehmann [2003]). For example, for

172   harem polygamy, where successful males generally mate with most or all of the females in their

173   harem, and the females generally mate with only one male, the effective size, for a Poisson

174   distribution of progeny number, is better approximated by

175    $$N_e = \frac{4 N_m N_f}{2 N_m + N_f},$$ (7)

176   rather than by eqn (4), showing the larger impact of male number for this type of mating system.

177   Other predictions of N_e for different mating systems and overlapping generations have been

178   provided by Nunney (1993).

179   Expression (5) can also be obtained following the concept of long-term contributions from

180   ancestors to descendants developed by Wray and Thompson (1990) in the context of populations

181   under selection. As suggested by Woolliams and Thompson (1994) and shown by Caballero and

182   Toro (2000, 2002), the expressions can be approximated by

183    $$N_e \approx \frac{2N}{(1 + V_\infty)(1 - \alpha)},$$ (8)

184   where V_∞ is the variance of long-term contributions from ancestors to descendants. For random

185   mating ($\alpha = 0$), $V_\infty = 1$ and $N_e = N$, as expected.

186   For a proportion β of partial selfing, $V_\infty \approx S_k^2 / [2(1 - \beta)]$ if α is not too close to one

187   (Caballero and Toro, 2000), which, when substituted into (8), gives eqn (5). When the numbers of

188   selfed and nonselfed progeny are independently Poisson distributed, $S_k^2 \approx 2 + 2\beta$ (see Nomura,

189   1999a for a more precise prediction), and both eqn (5) and (8) reduce to

190    $$N_e = \frac{N}{1 + \alpha}$$ (9)

191 (Li, 1976, p. 562; Pollak, 1987; Caballero and Hill, 1992; Nordborg and Krone, 2002). For a

192 population with Poisson distribution of family size and complete selfing, (9) reduces to $N_e = N/2$.

193 Equation (8) can also be applied to the scenario of partial full-sib mating with the appropriate

194 approximation for V_∞ (Caballero and Toro, 2000).

195   Predictions of the effective size for populations with mixed sexual and asexual reproduction

196 systems and discrete and overlapping generations have been developed by Yonezawa (1997).

197 Assuming a monoecious plant species carrying out asexual propagation with a proportion δ in a

198 population of constant size N (i.e. an average progeny number of one and two for asexual

199 propagation and sexual reproduction, respectively), eqn (5) of Yonezawa (1997) can be rearranged

200 to

201 $$N_e = \frac{4N}{[2(1-\alpha)+S_k^2(1+\alpha)](1-\delta)+4\delta S_c^2(1+\alpha)} , \qquad (10)$$

202 where S_c^2 is the variance of the number of asexually produced progeny among plants. If there is no

203 asexual reproduction ($\delta = 0$), eqn (10) reduces to (5), as it should. If all reproduction is asexual (δ

204 $= 1$), (10) reduces to

205 $$N_e = \frac{N}{S_c^2(1+\alpha)} . \qquad (11)$$

206 Interestingly, if the number of asexually produced progeny is Poisson distributed ($S_c^2 = 1$), the

207 expression is the same as for a sexually reproducing partially selfed population where the numbers

208 of selfed and nonselfed progeny are independently Poisson distributed, i.e. eqn (9). If all individuals

209 are homozygotes ($\alpha = 1$), $N_e = N/2$, the same as for a fully selfed population.

210   An extension of eqn (10) to overlapping generations was also given by Yonezawa (1997).

211 Equation (10) assumes that sexual and asexual contributions are independent. Predictions relaxing

212 this assumption and extensions to more complex models were given by Yonezawa *et al*. (2000,

213 2004). Analytical expressions for these models of mixed sexual and asexual species were also given

214 by Orive (1993) and Balloux *et al*. (2003) using coalescence theory.

215   Predictions of effective size for haplo-diploid species can generally be made by the standard

216 formula for sex-linked genes (see review by Caballero, 1994). Some situations occur, however,

217 where reproduction of these species is more complex than assumed by the simplest models. For

218 example, in many eusocial Hymenoptera species, males can be produced by workers rather than

219 only by queens. Predictions of N_e for this scenario have been developed by Nomura and Takahashi

220 (2012).

221

222 ***Populations under selection***

223

224 In the absence of selection or when selection acts on a non-inherited trait, the effective size is

225 simply a function of the variance of the number of offspring per parent, as in eqn (5). However,

226 predictions of N_e are more complicated when selection acts on an inherited trait, such as when

227 artificial selection is carried out for a quantitative trait in animal or plant breeding, or when natural

228 selection acts on fitness traits against deleterious mutations or in favour of advantageous ones. In

229 these scenarios, the drift process is amplified over generations because the random associations

230 originated in a given generation between neutral and selected genes remain in descendants for a

231 number of generations until they are eliminated by segregation and recombination. This problem

232 was first addressed by Robertson (1961) and later on by other authors (e.g., Wray and Thompson,

233 1990; Woolliams *et al.*, 1993; Santiago and Caballero, 1995) for directional selection in quantitative

234 traits. Extensions of the model were made later for populations under natural selection, linkage,

235 overlapping generations and animal breeding schemes, as will be reviewed below.

236

237 *Selection assuming unlinked genes*: When selection acts on an inherited trait, changes in gene

238 frequency at a focal neutral locus are positively correlated over generations because the selective

239 values randomly associated with the neutral locus are not completely removed by segregation and

240 recombination from one generation to the next. For unlinked genes and weak selection, the random

241 association generated by sampling in a single generation is halved in consecutive generations by

242 segregation and recombination. Therefore, the accumulative selective association has a limiting

243 value $Q = \sum_{i=0}^{\infty}(1/2)^i = 2$ times the value of the original random association (Robertson, 1961),

244 and the corresponding variance of the long-term contributions of copies of the neutral gene will

245 increase by a factor Q^2. With regards to drift, the effective variance of contributions of individuals

246 (with average 2) increases due to selection by the same factor up to $4Q^2C^2$, where the term C^2 is the

247 genetic variance of the individual trait measures (for the quantitative trait subject to artificial

248 selection, or fitness-related traits in the case of natural selection) relative to the mean of the trait in

249 the population. This variance has to be added to the expected variance of random contributions not

250 caused by selection (S_k^2) to predict the total variance of contributions. In reality, the associations are

251 also reduced each generation to a proportion equal to the fraction of genetic variance remaining

252 after selection (G) which, in turn, can be increased by the correlation between the selective

253 advantages of male and female parents (r), and the series becomes

254 $Q = \sum_{i=0}^{\infty}[G(1+r)/2]^i = \frac{2}{2-G[1+r]}$ (12)

255 (Santiago and Caballero, 1995). In the case of partial selfing (or partial full-sib mating), the term r

256 in eqn (12) should be replaced by β (the proportion of inbred matings), because the correlation

257 between the expected selective values of males and females (r) is approximately one for inbred

258   matings (which take place with proportion β) and approximately zero for non-inbred matings, i.e.

259   $Q = 2/[2 - G(1 + \beta)]$.

260   Therefore, the equation accounting for selection as an extension of eqn (5) is

261   $N_e = \dfrac{4N}{2(1-\alpha)+(S_k^2+4Q^2C^2)(1+\alpha)}$.   (13)

262   Nomura (1999b, 2005a) showed that eqn (13), obtained by a genetic drift approach, could

263   also be derived from an inbreeding approach by considering the variance of long-term contributions

264   as used by Wray and Thompson (1990) and Wray *et al*. (1990), when appropriate corrections are

265   made in the latter (see also Woolliams and Bijma, 2000).

266   The application to different numbers of males and females was given by Santiago and

267   Caballero (1995). That equation, however, lacked the same covariances as the equation without

268   selection, as shown by Nomura (1997a) and Wang (1998). For random mating ($\alpha = 0$) and Poisson

269   distribution of family sizes ($S_k^2 = 2$), eqn (13) reduces to the simplest expression (Robertson, 1961),

270   $N_e = \dfrac{N}{1+Q^2C^2}$.   (14)

271   Equation (14) can be expressed in terms of heritability (h^2) of fertility, as shown by Nei and

272   Murata (1966) and Nomura (2002b). Let V_k be the observed variance of family sizes, which would

273   be $V_k = S_k^2 + 4C^2$ if the decay in the cumulative effect of selection is ignored (i.e., $Q = 2$). The first

274   term, $S_k^2 = V_k(1 - h^2)$, is the non-heritable component of this variance, and the second term,

275   $4C^2 = V_k h^2$, is the heritable component. Thus, substituting these into eqn (14) yields

276   $N_e = \dfrac{4N}{2+(1+3h^2)V_k}$ (15)

277   (Nei and Murata, 1966). The extension of eqn (15) to dioecious populations was developed by Nei

278   and Murata (1966) assuming random union of gametes. A more general equation was developed by

279   Nomura (2002b), who also suggested a form of the equation that avoids estimating the heritability,

280   $N_e = \dfrac{4N}{2+V_k+3\left(cov_{k,m}+cov_{k,f}\right)}$,   (16)

281   where $cov_{k,f}$ and $cov_{k,m}$ are the offspring-mother and offspring-father covariances of sibship size,

282   respectively.

283   The prediction of effective population size under selection with overlapping generations was

284   considered by Nomura (1996) and Bijma *et al*. (2000). As for the non-selection case (Hill, 1979), N_e

285   is the same as that for populations with discrete generations having the same non-selective and

286   selective components of variance in lifetime progeny numbers, and the same number of individuals

287   entering the population each generation. Another interesting result is that the average age of parents

288   in populations under selection is smaller than that in populations under no selection, as, in the

289   former, younger parents tend to have higher selective advantages.

Genetic marker data can be used to assist selection (i.e. marker assisted selection, MAS) for a quantitative trait. Nomura (2000) investigated the predictive equation of N_e in this scenario and showed that it depends on the relative values of the genetic (r, 0.5 for full-sib families) and phenotypic (t) correlation between family members, where $t \approx h^2/2 + c^2$, h^2 is the trait heritability and c^2 is the fraction of the phenotypic variance due to the shared common environments of family members. When an index (I) is considered using individual phenotype (P) and molecular marker (M) information with given weights, i.e. $I = B_P P + B_M M$, the effective size with MAS is reduced relative to that for phenotypic selection alone (eqn [14]) when $t < r$, and is increased when $t > r$.

The prediction of effective population size under index selection was addressed by Wray *et al*. (1994), Caballero *et al*. (1996b) and Nomura (1998b, 2005a). Assume truncation selection is carried out based on an index selection of the individual phenotype (P) and the mean phenotype of its full-sib family (P_f, including the individual), $I = B_w (P - P_f) + B_b (P_f)$, where B_w and B_b are the corresponding selection weights. The effective size can then be predicted using eqn (14), where $Q = 2/(1 + k B_b)$ and $C^2 = S_k^2 + 4i^2(\rho_I - \rho_A) + 4i^2 \rho_A$, where ρ_I is the correlation of full sibs for the index values, ρ_A is the correlation of full sibs due to the breeding value of the parents, $k = i (i - x)$, i is the selection intensity and x is the truncation point in the standardised normal distribution. This predictive equation corrects a typographical error in a sign in the equation of Caballero *et al*. (1996b, p.77). When the whole pedigree information is available, estimation of breeding values can be made by Best Linear Unbiased Prediction (BLUP) selection. Predictions of the effective size under this selection method were investigated by Nomura *et al*. (1999), Bijma and Woolliams (2000), and Bijma *et al*. (2001).

Other extensions for the prediction of the effective population size under selection refer to sex-linked loci (Nomura, 1997b; Wang, 1998), gynodioecious species, (i.e. species which have both hermaphrodite and female individuals, Laporte *et al*., 2000), open nucleus schemes (Nomura, 1997c; Bijma and Woolliams, 1999), and selection on traits affected by maternal effects (Rönnegård and Woolliams, 2003).

Selection at linked loci: The above formulations predict the rates of inbreeding that are usually calculated by tracing paths in genealogies of individuals. However, the real rates of inbreeding are expected to be larger than those predictions when selection acts on a system of linked genes. The reason for this is that the two gene copies at a neutral locus in an individual have different probabilities of propagation to the next generation, because they are embedded in homologous chromosomes with different alleles at linked selected loci. The problem of predicting N_e in populations under purifying selection with linkage (the background selection model; Charlesworth, 2013) was addressed by Hudson and Kaplan (1995) and Nordborg *et al*. (1996) focusing on the

325   effect of selection on nucleotide diversity, by Santiago and Caballero (1998) analysing the

326   cumulative effect of selection over generations with a genetic drift approach (the Robertson effect),

327   and by Nicolaisen and Desai (2012, 2013) using the coalescence theory. All these papers derived

328   the same equation for the asymptotic N_e, which is a function of the haploid deleterious mutation rate

329   U, the effect s of mutations and the length L of the whole-genome or genome segment given in

330   Morgans,

331    $N_e = N e^{-U/(s + L/2)}$.$\hspace{8cm}$(17)

332   This equation is the exponential version of eqn (14), $N_e = N e^{-Q^2 C^2}$, which was derived from the

333   multiplicative fitness model assumed under background selection. Here, $C^2 = Us$ is the variance for

334   fitness and the cumulative term, Q, for a rate of recombination c between the neutral and selected

335   loci, is $Q_c = \sum_{i=0}^{\infty}[(1-s)(1-c)]^i \approx 1/(s+c)$ (Santiago and Caballero, 1998). If the focal

336   neutral locus is located in the middle of the genome segment and the selected loci are uniformly

337   scattered, the average value of the Q_c^2 terms over the segment is $Q^2 = 1/[s(s + L/2)]$. Substituting

338   this and $C^2 = Us$ in eqn (14), we obtain eqn (17).

339   It is important to note that this equation predicts the magnitude of drift or inbreeding in the

340   long-term. For the focal neutral allele, this magnitude is effectively reached after a number of

341   generations counted since it first appeared by mutation. Until that moment, drift at the neutral locus

342   is expected to increase with time. The increasing drift acting on neutral mutations in consecutive

343   generations can be predicted by the partial $N_{e(t)}$ values for generation t forward in time that can be

344   calculated using the partial cumulative terms $Q_{c(t)} = \sum_{i=0}^{t}[(1-s)(1-c)]^i = (1 - e^{-(s+c)t})/(s +$

345    $c)$ (Santiago and Caballero, 1998). An equivalent conclusion was reached by Nicolaisen and Desai

346   (2012, 2013) from the point of view of the coalescent process. The consecutive $N_{e(-t)}$ values that

347   predict the increasing probability of coalescence under selection t generations backwards in time

348   (thus, the negative sign) reach an asymptotic value given by eqn (17) and the predictions of the

349   partial N_e values given by both methods, forward and backward, are exactly the same for any

350   generation t,

351    $N_{e(t)} = N_{e(-t)} = N e^{-\frac{Us}{L/2}\int_0^{L/2} Q_{c(t)}^2 dc}$.$\hspace{6cm}$(18)

352   Illustrations of the decline in $N_{e(t)}$ over generations are given in Fig. 3 of Santiago and Caballero

353   (1998) and Fig. 3 of Nicolaisen and Desai (2013). This shows that the distortion of coalescent

354   genealogies under selection and the cumulative effect of selection over generations are both

355   specular images of the same process. Moreover, the pattern of neutral variation in populations under

356   selection can be predicted by accumulating the expected distributions of neutral mutations

357   originated in all the previous generations with the corresponding consecutive effective sizes given

358   by the $N_{e(t)}$ values (Santiago and Caballero, 1998; Nicolaisen and Desai, 2012, 2013). This means

359   that the intensity of genetic drift varies over mutations that occurred at different generations

360   backward in time. Therefore, strictly speaking, there is not a single N_e value representing the

361   intensity of the drift process and, consequently, the amount and spectrum of variation under

362   selection, even in populations at mutation-selection-drift equilibrium.

363   If mutations are advantageous (selective sweep model), predictions are generally

364   cumbersome, because the genetic variance at selected loci is a function of the gene frequencies.

365   Moreover, the stochastic distribution of selective sweeps over time is far away from the constant

366   flux of variation usually assumed to simplify derivations. Wiehe and Stephan (1993) and Gillespie

367   (2000) derived equations for the expected heterozygosity at neutral loci using a model in which

368   recurrent favourable mutations pass quickly through the population to fixation, wiping out linked

369   variation. The first key simplification in these derivations is that the time of fixation of favourable

370   mutations is so short relative to the time scale of genetic drift that it can be considered as occurring

371   instantaneously. The second simplification is that the rate of recovery of neutral variation after a

372   selective sweep is very slow when compared with the rate of occurrence of the sweeps. The

373   recurrent substitutions lead to a roughly constant level of neutral variation in a stochastic process

374   that is often referred to as "genetic draft". A simple solution for N_e can be obtained by directly

375   applying the concept of variance of long-term contributions to an evolutionary time-scale:

376   $$N_e = \frac{N}{1+2N\theta y^2}$$ (19)

377   (Gillespie, 2000), where θ is the rate of selective sweeps and y is the final frequency of the neutral

378   copy that was initially associated with the favourable mutation when it first appeared. The

379   frequency of this neutral copy is expected to increase to Ny copies after the sweep, and the

380   frequency of each of the other $2N - 1$ copies is expected to be reduced to $(1 - y)$ copies. Therefore,

381   the variance of the expected long-term contributions for a single selective sweep is about $2Ny^2$. As

382   selective sweeps occur at a rate θ, the second term of the denominator in equation (19) is the total

383   variance of the expected long-term contributions, that is Q^2C^2 in equation (14).

384

385   *Effective population size in conservation practices*

386

387   The concept of effective size is key to conservation genetics practices, as it summarises the past

388   history of the population regarding inbreeding and genetic drift, and provides the prospects for the

389   sustainability of the population if the current effective size is maintained in the future. The effective

390   population size is directly related to the statistics widely used to monitor conservation breeding

391   schemes, such as the number of genome equivalents $N_{ge} \approx N_e/t$ (Lacy, 1995), where t is the

392   number of generations of management.

393   Minimising the loss of genetic variation is one of the main objectives of captive breeding

394   programmes. This is achieved through minimising genetic drift and, therefore, maximising N_e. A

395   classical strategy to follow is the equalisation of family sizes. By choosing one couple from each

396   pair of parents, the variance of parental contributions is null ($S_k^2 = 0$) and, from eqn (1), $N_e \approx 2N$

397   (Wright, 1938; Crow, 1954), which is twice the effective size of an unmanaged population with a

398   Poisson distribution of family size. This is known as minimal inbreeding and it is the recommended

399   procedure for applications in germplasm collection and regeneration in plants (see, e.g., Vencovsky

400   *et al*., 2012). However, effective sizes larger than $2N$ can actually be obtained by population

401   subdivision (Wright, 1943; Wang and Caballero, 1999) and other strategies, as shown below. The

402   extension of the strategy in the case of different numbers of males and females was developed by

403   Gowe *et al*. (1959). In their selection scheme, each male contributes one son and $r_{fm} = N_f/N_m$

404   daughters, and each female contributes one daughter and has a probability of N_m/N_f of contributing

405   one son. Thus, $S_{fm}^2 = (N_m/N_f)[1 - (N_m/N_f)]$ and all of the other variances and covariances in

406   eqn (2) are zero. Substituting these into eqns (2-3) gives

407    $N_e = \frac{16N_mN_f}{3N_f+N_m}$.   (20)

408   Compared with no selection (random Poisson distribution of the number of offspring per parent,

409   eqn 4), this scheme can increase N_e by a proportion of $(r_{fm} + 3)/(3r_{fm} + 1)$. When the

410   female/male ratio, $r_{fm} = N_f/N_m$, is 2 , for example, N_e is increased by 71.4%.

411   Wang (1997c) proposed an alternative design which produces further increases in N_e of

412   about 17% when $r_{fm} = 2$. In this scheme, among the r_{fm} females mated with each male, one is

413   selected at random to contribute one son, and each of the remaining $r_{fm} - 1$ females contributes one

414   daughter. In this scenario, S_{fm}^2 is as for eqn (20), but $S_{ff}^2 = 2N_m/N_f$, and a negative covariance is

415   induced between the numbers of male and female offspring from female parents, $S_{fm,ff} =$

416    $- N_m/N_f$. Substituting these terms into eqns (2-3),

417    $N_e = \frac{16N_mN_f^2}{3N_f^2+2N_m^2-N_mN_f}$.   (21)

418   The benefit of this scheme over that from Gowe *et al*. (1959) is decreased as r_{fm} gets larger. For sex-

419   linked loci a benefit is also produced if males are the heterogametic sex. The above equations refer

420   to random mating of parents. Wang (1997c) also proposed a system of non-random mating in which

421   each male is mated with one of the groups of half-sib females that are not sisters of the male. This is

422   a sort of population subdivision where the half sibs are like "subpopulations" and there is random

423   migration of males and no migration of females among the "subpopulations" (see Wang and

424   Caballero, 1999). The mating scheme can further increase N_e over that predicted by eqn (21).

425   Wang´s (1997c) method applies to a single generation. Sánchez *et al*. (2003) extended the

426   method across generations to account for long-term contributions, further improving its efficiency.

427   With the availability of pedigree or molecular marker information, a more general method, based on

428   finding the contributions from parents to progeny which minimise the average coancestry among

429   the progeny (minimum coancestry contributions), is the most widely proposed criterion to maintain

430   genetic diversity (Toro and Pérez-Enciso, 1990; Ballou and Lacy, 1995; Fernández *et al*., 2003;

431   Meuwissen, 2007). This method has been shown to minimise the variance of the long-term

432   contributions from ancestors to descendants and, therefore, to maximise effective population size

433   (Caballero and Toro 2000, 2002).

434   The above methods are all designed to reduce the variation in family sizes, the term S_k^2 in

435   eqn (1) and its corresponding components when the numbers of males and females are different

436   (eqn 2). It is also possible to increase the effective population size by decreasing the Mendelian

437   segregation variance, which is represented by the *constant* term "2" in eqn (1). This latter can be

438   achieved by the use of Marker Assisted Selection (MAS) to minimise the variation in contribution

439   between the paternally and maternally derived genes at a locus (Wang and Hill, 2000). Thus, for

440   example, for equal numbers of males and females and equalisation of individual contributions, N_e

441   can be expressed as

442    $$N_e = \frac{2N}{P_{m,mf} + P_{f,mf}},$$ (22)

443   where $P_{m,mf}$ ($P_{f,mf}$) is the probability that the two genes coming from the male (female) parent and

444   contributing to their male and female progeny are copies of the same gene. By MAS it is possible to

445   reduce these probabilities below the value of 0.5 expected under no control of Mendelian

446   segregation, depending on the amount of marker information, the genome size and the number of

447   marker-genotyped offspring per family, achieving values of N_e larger than 2N. MAS can also be

448   used in a more general framework of different numbers of males and females to minimise global

449   genetic drift and inbreeding (Wang, 2001a).

450   An alternative and complementary method is to use reproductive technologies for meiosis

451   manipulation, such as *in vitro* culture of premeiotic germ cells and microinjection of primary

452   spermatocytes into oocytes. By using more than one gamete from a single meiosis, variation from

453   Mendelian segregation can be partially or completely removed (Santiago and Caballero, 2001).

454   Thus, for example, if equalisation of family sizes is carried out and the gametes from both male and

455   female parents are managed to come from the same meiosis in each case, the resulting effective size

456   becomes 3N, rather than the typical 2N.

457   The control of the increase in inbreeding and genetic drift in conservation programmes is

458   mainly addressed by reducing the variances of genetic contributions between paternally and

maternally derived genes within and between individuals by equalizing family sizes and minimizing

Mendelian segregation variance, as discussed above. A minor contribution to this control can also

be achieved by the avoidance of inbred matings and other types of non-random mating. The simple

avoidance of sib mating has a very minor effect (Wang, 1997d) and methods such as the maximum

avoidance of inbreeding (MAI) proposed by Wright (1921) have a higher, although still relatively

low, impact. These can be carried out after the design of parental contributions has been

implemented (Caballero *et al*., 1996a; Meuwissen, 2007). Alternatively, avoidance of inbreeding

and optimal parental contributions can be realized in a single step (Fernández *et al*., 2004) by the

so-called mate selection method. Systems of mating involving circular (half-sib mating) (Kimura

and Crow, 1963b; Theodorou and Couvet, 2010) or rotational schemes (Nomura and Yonezawa,

1996) generally reduce the ultimate rate of inbreeding, but at the cost of higher initial rates

(Robertson, 1964), so that their use in conservation is not recommended because the high risk of

extinction from inbreeding depression. Some methods are of particular application in selection

programmes, such as the compensatory mating proposed by Santiago and Caballero (1995), where

individuals from large families are mated to individuals from small ones. This produces negative

correlations between the drift caused by selection and the drift caused by sampling, partly

counteracting the cumulative effect of selection represented by the term Q in eqn (12). This system

of mating in combination with MAI allows for a substantial reduction of inbreeding (Caballero *et

al*., 1996a).

A conservation strategy of high relevance in fisheries is supportive breeding (Hare *et al*.

2011), where a population is typically divided into a captive and a wild group and the offspring of

the captive group are released into the wild habitat to mix with the offspring from the wild group.

Because the captive group (permanent or transitional) is bred to produce a lot of offspring that are

released into the wild group at each generation, the variance in family size is greatly elevated

artificially and thus the N_e of the entire population is reduced. Ironically, the more successful the

supportive program is in augmenting the wild population, the greater the reduction in N_e and the

greater the loss of genetic diversity in the total population (supportive + wild). This paradox is

overcome only when successful supportive breeding in augmenting the wild population is carried

out over a long period of time such that the excessive drift and inbreeding in the initial generations

of supportive breeding is compensated for by weaker drift and inbreeding in later generations

because of the increase in census size. Ryman and Laikre (1991), Ryman *et al*. (1995) and Wang

and Ryman (2001) have provided approximations for the inbreeding and variance effective sizes,

respectively, which can be different in this case, with one generation of supportive breeding.

Nomura (1998a) obtained an expression of N_e from the change in coancestry, which agrees with the

493   variance effective size, as expected. In this scenario, with various census sizes and a mixture of

494   groups, predictions depend, however, on the generations considered (see Ryman *et al*. 1999).

495   For endangered species in the wild, estimating the effective population size and monitoring

496   its changes over time is important in understanding the genetic health, evaluating the risk of

497   inbreeding and inbreeding depression and thus the risk of extinction, assessing the effectiveness of

498   the genetic managements (e.g. human-aided migration/relocation, habitat protection or

499   modification), and projecting the future demographic trajectories of the populations. Simulations

500   (Tallmon *et al*. 2010) showed that monitoring the effective size is most often a more robust means

501   of identifying stable and declining populations than monitoring census size. If a population is

502   detected to have a small or declining N_e, the managers of the populations should be alerted to

503   investigate the causes and to take effective measures for reversing the course. Using noninvasive

504   sampling (e.g. from faeces, feathers, hair, …), genetic marker data can be obtained from a

505   population even without observing (disturbing) the animals. The data can then be analysed

506   normally, except for accounting for the possibility of genotyping errors and allowing for a high

507   frequency of missing data (e.g. Wang 2004), in estimating N_e. For example, in a long-term

508   monitoring project, Kamath *et al*. (2016) sampled and genotyped (at 20 microsatellite loci) 729

509   Yellowstone grizzly bears (*Ursus arctos*) born in the period 1962-2010 from an isolated and well-

510   studied population in the Greater Yellowstone Ecosystem. They used the data to study the

511   population demographic trajectories, estimating the changes in N_e and generation interval, over this

512   time period.

513

514 Methods for estimating the effective population size from genetic data

515

516   Given the concepts of effective size, different approaches can be used to *predict* the effective size of

517   a population from its demographic parameters such as census size and variance of reproductive

518   success. In parallel, different methodologies can also be developed to *estimate* the realized effective

519   size of a population from its genetic properties revealed by genetic markers, such as temporal

520   changes in allele frequency and linkage disequilibrium.

521   Quite a few methods (Schwartz *et al*., 1999; Beaumont, 2003a; Wang, 2005; Palstra and

522   Ruzzante, 2008; Luikart *et al*., 2010; Gilbert and Whitlock, 2015) have been developed and applied

523   to estimating N_e in widely different spatial and time scales, from ancient, past to current (parental)

524   population sizes. Herein we focus on the effective size of the current generation or just a few

525   generations in the past, as this time scale is the most relevant for conservation genetics (Luikart *et*

526 *al.*, 2010) and plant and animal breeding, and is most likely to yield accurate estimates in current

527 practices.

528

529 ***Heterozygosity excess***

530

531 Compared with an infinitely large population at Hardy-Weinberg equilibrium, a population

532 generated from a number of N_m male parents and a number of N_f female parents is expected to show

533 a deficit of homozygotes and an excess of heterozygotes at a neutral locus when N_m, N_f, or both, are

534 small. This is because male and female parents are expected to have different allele frequencies due

535 to drift. The smaller the value of N_m or N_f, the greater the difference between paternal and maternal

536 allele frequencies and thus the greater the excess in heterozygosity of the offspring population.

537 There is a simple functional relationship between the N_e of the parental population and the amount

538 of heterozygosity excess in the offspring population (e.g. Robertson, 1965; Wang, 1996a). For a

539 Wright-Fisher ideal population except for separate sexes with N_m male and N_f female parents, the

540 heterozygosity excess is expected to be

541 $$D \approx -\frac{1}{8N_m} - \frac{1}{8N_f} = -\frac{1}{2N_e}, \tag{23}$$

542 where $N_e = 4N_m N_f / (N_m + N_f)$ is the effective size of the parental population given by eqn (4).

543 The value of D is negative, indicating an excess of heterozygosity and a corresponding deficit of

544 homozygosity. For a non-ideal population with arbitrary distributions of family sizes, (23) is still

545 valid when N_m and N_f are replaced by N_{em} and N_{ef} respectively, the effective numbers of male and

546 female breeders.

547 Eqn (23) suggests that measuring the heterozygosity excess, D, at a number of marker loci in

548 a population yields an estimate of the parental population effective size. Pudovkin *et al.* (1996)

549 proposed such a N_e estimator by accounting for the sampling effect,

550 $$\widehat{N}_e = \frac{1}{2\widehat{D}} + \frac{1}{2(\widehat{D}+1)}, \tag{24}$$

551 where the observed heterozygosity excess is estimated by $\widehat{D} = \widehat{H}_e / (\widehat{H}_e - \widehat{H}_o)$, $\widehat{H}_e = 2\hat{p}(1 - \hat{p})$ is

552 the expected heterozygosity from the observed gene frequency $\hat{p}$, and $\widehat{H}_o$ is the observed

553 heterozygosity. $\widehat{D}$ is calculated for each allele in a multiallelic locus and for each locus, and the

554 average value is used in (24) (Luikart and Cornuet, 1999). The accuracy of the estimator was

555 evaluated by Pudovkin *et al.* (1996) using simulations, and was applied to a few empirical datasets

556 (Luikart and Cornuet, 1999). The method is simple, and is implemented in several computer

557 programs (e.g. Zhdanova and Pudovkin, 2008; Jones and Wang, 2010; Do *et al*. 2014). However,

558 the method has a low precision and accuracy, frequently providing infinitely large estimates of N_e

559 for small populations. The estimator is also highly sensitive to non-random mating (e.g. population

18

560 subdivision, close relative mating), which also causes deviation from Hardy-Weinberg equilibrium.

561 Its poor performance renders it useless in applications to empirical dataset analysis, except when the

562 actual population size is very small and marker information is ample.

563

564 *Linkage disequilibrium*

565

566 In a large unselected random mating population, alleles are independent within and between loci,

567 producing Hardy-Weinberg equilibrium and linkage equilibrium. In a finite population, however,

568 random genetic drift leads to associations between alleles at a locus and between alleles of different

569 loci. The former results in heterozygosity excess, and the latter leads to gametic linkage

570 disequilibrium (LD). In addition to drift, LD can also be induced by factors such as migration and

571 direct or indirect (e.g. hitchhiking) selection (Hedrick *et al*., 1978). For neutral loci unlinked with

572 selected loci in an isolated population under random mating, LD would come exclusively from

573 genetic drift and can be used to estimate N_e (Hill, 1981).

574   A LD estimator of N_e for a random mating population at equilibrium is based on the

575 formulation (Hill, 1981),

576 $$E[r^2] = V[r] = \frac{(1-c)^2 + c^2}{2N_e c(2-c)} + \frac{1}{n} \,, \qquad\qquad (25)$$

577 where c is the recombination rate ($c = 1/2$ for unlinked loci), r is the correlation of allele frequencies

578 between two loci due to LD, and n is the sample size (number of sampled individuals). In an

579 equilibrium population, allele frequencies at two neutral loci are expected to be uncorrelated (i.e.

580 $E[r] = 0$), such that the expectation of squared r, $E[r^2]$, is equal to the variance of r, $V[r]$. Eqn (25)

581 shows that $V[r]$ is composed of two distinctive parts. The first comes from genetic drift, determined

582 by N_e and linkage c. The second comes from sampling, determined by sample size n. Using the

583 genotypes at a number of loci of n sampled individuals, we can estimate $V[r]$, which can then be

584 inserted in (25) to obtain an estimate of N_e if the recombination fraction c between loci is known.

585 Note that a slightly different expression for the population $V[r]$ (i.e. the first part on the right-hand

586 side of eqn 25) was derived by Sved (1971) from an identity by descent approach. For a dioecious

587 population with monogamy, the right side of (25) should be increased by $c / (2N_e c(2 - c))$ (Weir and

588 Hill, 1980).

589   Hill (1981) also derived the formula for the sampling variance of the estimator such that

590 uncertainties of the N_e estimates can also be evaluated. For the case of no linkage, Waples (2006)

591 showed by simulations that the LD estimator can seriously underestimate N_e when sample size is

592 small. He derived empirical equations to correct for the bias caused by small sample sizes, and

593 showed by simulations that the accuracy of the modified estimator is comparable to the temporal

594   method described in the next section. To facilitate the applications of the LD estimator, Waples and

595   Do (2008) published a computer program, LDNE, and further evaluated its performance in

596   comparison with the temporal method, using simulated data (Waples and Do, 2010). They

597   concluded that, under similar conditions in terms of marker information and the actual population

598   size, LD estimator can yield N_e estimates that have equivalent or better qualities than the temporal

599   estimators, except when the sampling interval of the temporal method is long.

600   The LD estimator is simple to calculate, and requires just a single sample of multilocus

601   genotypes instead of two or more samples, as is with the temporal method (see below). It is

602   especially suitable for species with a long generation interval where obtaining two samples

603   separated by a couple of generations means many years, and for genetic monitoring (Schwartz *et*

604   *al*., 2007) to track population trajectories on a yearly basis. As a result, the LD estimator has gained

605   popularity in recent years (Palstra and Ruzzante, 2008; Luikart *et al*., 2010). However, some

606   assumptions inherent to the LD estimators are often violated in real populations, and as a result may

607   lead to biased N_e estimates. For example, it is assumed that LD is produced solely from the finite

608   population size, and other confounding factors, such as non-random mating and population

609   structure, are absent. Any departure from random mating (e.g. an excess or deficit of close relatives

610   mating including selfing) will affect LD, and thus LD based estimates of N_e. Waples *et al*. (2014)

611   evaluated the effect of age structure on LD estimators, and found that LD calculated from mixed-

612   age adult samples is overestimated and thus N_e is underestimated in all of 21 simulated species with

613   different life tables. Similarly, the LD in a subpopulation is reduced by a constant and high rate of

614   immigration and elevated by a low rate of immigration, compared with that of an isolated

615   population of the same N_e. Therefore, as observed by Waples and England (2011) in their

616   simulation study, LD calculated from a sample from a subpopulation leads to an overestimate or an

617   underestimate of local N_e when immigration rate is high or low, respectively. In the former case, the

618   estimated local N_e converges to the global N_e of the entire population (Waples and England, 2011).

619   LD is highly dependent on the recombination rate between loci (Hill, 1981). Pairs of closer

620   linked loci have higher LD, and thus provide better information about N_e (suitably defined in time

621   scale) if the linkage information among loci is known. Although Hill (1981) derived his LD

622   estimator of N_e allowing for an arbitrary level of linkage, and he advocated the use of linked

623   markers, most often unlinked markers are used in practice because either truly unlinked markers are

624   used or potentially linked markers are used but their linkage relationship is unknown. LD of

625   markers of different recombination rates sheds light on the effective size of the population in

626   different time periods in the past (Wang, 2005). Quite a few methods (Hayes *et al*. 2003; Barbato *et*

627   *al*. 2015; Mezzavilla and Ghirotto 2015; Saura *et al*. 2015) have been developed to exploit the LD

628 information from many densely spaced markers on a chromosome segment in inferring the N_e at

629 different time points in the past.

630

631 *Temporal changes in allele frequency*

632

633 For an infinitely large population under Hardy-Weinberg equilibrium, both allele frequencies and

634 genotype frequencies remain unchanged over time. In reality, these frequencies never stay constant,

635 and change systematically due to the forces of mutation, selection and migration, stochastically due

636 to the random force of genetic drift, or both. In the absence of the action of all of the systematic

637 forces in a population, any observed allele frequency change must come solely from genetic drift

638 and can thus be used to infer the rate of drift or the N_e of the population. Based on this logic,

639 Krimbas and Tsakas (1971) proposed to measure allele frequency changes at a number of marker

640 loci between two temporally separated samples of individuals, and thereby to estimate the N_e of the

641 population during the sampling interval. This so-called 'temporal method' was subsequently

642 developed by many others in both (allele frequency) moment (e.g. Nei and Tajima, 1981; Pollak,

643 1983; Waples, 1989) and likelihood (e.g. Williamson and Slatkin, 1999; Anderson *et al*., 2000;

644 Wang, 2001b; Berthier *et al*., 2002; Beaumont, 2003b; Laval *et al*., 2003) approaches.

645   Moment estimators calculate a standardized variance in the temporal changes of allele

646 frequency, F, from marker genotypes in two temporally spaced samples. F is essentially similar to

647 Wright's F_{ST}, the differences being that F measures the temporal differentiation for the same

648 population and it also includes sampling effect. There are a few F estimators (e.g. Nei and Tajima,

649 1981; Pollak, 1983) available, the one being widely applied was derived by Nei and Tajima (1981).

650 This estimator is calculated by

651 $$\hat{F} = \frac{1}{k}\sum_{i=1}^{k}\frac{(x_i-y_i)^2}{(x_i+y_i)/2-x_iy_i},\qquad\qquad(26)$$

652 for a locus with k alleles, where x_i and y_i are the observed frequencies of allele i in the first and

653 second samples respectively. For multiple loci, $\hat{F}$ is obtained by averaging single locus estimates.

654 The expectation of $\hat{F}$ depends on the sampling schemes (sampling with or without replacements),

655 and is a function of N_e and sample sizes to account for genetic drift and sampling effects. Solving

656 the expectation equation of $\hat{F}$ for N_e yields the temporal estimate of the (harmonic) mean N_e during

657 the sampling period (Nei and Tajima, 1981; Waples, 1989).

658   Moment estimators rely on the summary statistic, F, which is simple to calculate. However,

659 they do not use the full allele frequency information and are thus less accurate than the probabilistic

660 methods. The latter, likelihood or Bayesian, are much more complicated in statistical modelling and

661 in computation. In general, temporal methods provide good estimates of N_e when it is not large,

662   sampling interval is not too short (e.g. one generation), and the assumptions of the methods are

663   satisfied, using a typical set of 10-20 microsatellites. Likelihood methods generally have higher

664   accuracy and precision than moment methods, especially for markers with rare alleles, as verified

665   by several extensive simulations (e.g. Wang, 2001b; Berthier *et al.*, 2002; Tallmon *et al.*, 2004).

666   They are however much more computationally demanding than moment methods, which complete

667   an analysis almost instantly. Recently, the computational efficiency of likelihood methods has been

668   improved substantially by Hui and Burt (2015), using a hidden Markov algorithm and applying

669   continuous approximations to allele frequencies and transition probabilities. The new method can

670   deal with N_e values as high as several millions, and is implemented in an R package called NB.

671   A constraint on the applications of temporal approaches is the requirement of at least two

672   samples taken at one or preferably more generations apart. The longer this sampling interval, t, is,

673   the stronger the drift signal will be in the temporal data, and the more accurate the N_e estimate will

674   become. The extent of drift is proportional to t and is inversely proportional to N_e. For the same

675   population and the same sampling intensity (in terms of the number of markers, number of samples,

676   and number of individuals per sample), increasing the sampling interval t could improve the

677   accuracy of the temporal methods tremendously, as has been repeatedly verified by simulations

678   (e.g. Nei and Tajima, 1981). Moment estimators have especially low accuracy when t is small (say,

679   $t < 3$) because of the weak drift signal and also because of the approximations made to the

680   estimators. In practice, it is unfortunately difficult or unrealistic to increase t, especially for long-

681   lived species having a long generation interval.

682   Compared with other N_e estimating approaches, the temporal approach makes fewer

683   assumptions and is more robust to some complications (realities) in real populations. For example,

684   the approach is robust to population structure. It applies to a single unsubdivided population under

685   non-random mating (including selfing), and to a subdivided population when sampling is

686   representative and the aim is to estimate the N_e of the entire subdivided population rather than that

687   of a single subpopulation. It is also robust to age structure in a population with overlapping

688   generations, so long as the sampling interval t is large (e.g. Nei and Tajima, 1981). When t is small,

689   however, Waples and Yokota (2007) showed by simulations that typical sampling regimes

690   (sampling only newborns, only adults, and all age classes in proportions) result in biased N_e

691   estimates. Jorde and Ryman (1995) developed a moment estimator of N_e applicable to populations

692   with overlapping generations. They derived an age-structure correction factor, which, when applied

693   to the standard moment estimator for populations with discrete generations, leads to unbiased N_e

694   estimates for populations with overlapping generations. Unfortunately, however, the correction

695   factor is a function of numerous age-specific survival rates and age-specific reproduction rates of

696   the focal population. These rates are usually unknown. In fact, once all these rates are known for a

697 population, the N_e of the population can be calculated from standard N_e prediction equations (e.g.

698 Felsenstein, 1971; Hill, 1972, 1979) without the need of genetic data. A method to calculate the N_e

699 of a population with overlapping generations from its demographic parameters has been

700 implemented in an R-package Neff (Grimm *et al*. 2016).

701 The standard temporal approach for a single unsubdivided population was also extended to

702 estimate the N_e of a subpopulation which is connected to other subpopulations by gene flow (Wang

703 and Whitlock, 2003). Although both drift and immigration change allele frequencies of a

704 subpopulation, the detailed patterns of the changes are different between drift and immigration.

705 Using temporal samples from a focal subpopulation and a sample from a large source population

706 (the island-mainland model) or from two focal subpopulations (the island-island model), a moment

707 estimator and a likelihood estimator can yield joint estimates of N_e and migration rates, *m*.

708 Simulations showed (Wang and Whitlock, 2003) that both moment and likelihood estimators gave

709 reasonably good estimates of N_e and *m* under typical sampling intensities. However, no estimators

710 are available for the more general case of multiple ($n > 2$) subpopulations. Part of the difficulty is

711 with the number, n^2, of parameters to be jointly estimated, including *n* effective sizes and $n(n-1)$

712 migration rates. More work is badly needed in this direction as spatial and temporal genotype data

713 are becoming easy to collect thanks to the rapid developments in molecular technologies.

714 One of the assumptions in the temporal approach is the absence of selection so that any

715 change in allele frequency comes solely from drift and thus indicates the effective size of the

716 population. For most marker loci, the assumption is valid, especially for a small population over a

717 short sampling interval of just a few generations. However, over a long period, some loci could be

718 affected by adaptive selection or purging selection and their allele frequencies could change faster

719 or slower than those of neutral loci unaffected by selection. Allele frequencies at neutral loci could

720 also evolve faster or slower because of linkage disequilibrium with those under selection. The

721 temporal methods have been extended to estimate N_e of a population and the selection coefficient, *s*,

722 of a locus from time-series data of allele frequencies (e.g. Bollback *et al*., 2008; Mathieson

723 and McVean, 2013; Foll *et al*., 2015). These methods are usually Bayesian, based on

724 hidden Markov models to explain the observed allele frequency changes due to drift and selection.

725 How well these methods perform has yet to be checked, perhaps by a simulation study.

726

727 *Relatedness and relationship*

728

729 The pattern of genetic relatedness or relationship between individuals in a population has a direct

730 functional relationship with the inbreeding effective size of the population (Wang, 2009). Two

731 individuals taken at random from a population with a smaller N_e will have a higher probability of

732   sharing the same father, mother, or both. More generally, the mean and variance in pairwise

733   relatedness within a generation are expected to increase with decreasing N_e. Based on this logic,

734   Nomura (2008) proposed a method to use the increase in average coancestry between two

735   consecutive generations to estimate N_e. He showed by simulations that his coancestry method is

736   more biased, but more precise than the heterozygosity excess method. The overall accuracy

737   (measured by mean squared errors) of the two methods is similar. A major problem which causes

738   the bias of the method, as recognized by Nomura (2008), is that some non-sib pairs must be selected

739   from a sample of individuals to act as reference in estimating the mean coancestry. The selection of

740   non-sib pairs is difficult and somewhat subjective, because it is now well-known that classifying

741   dyads into even well separated relationship categories, e.g. full sibs, half sibs, parent offspring, and

742   unrelated, from pairwise relatedness estimates is highly error prone (e.g. Blouin *et al.*, 1996).

743   Although many marker-based pairwise relatedness estimators are unbiased, they have high

744   sampling errors with no exceptions (Wang, 2014).

745   A more robust and powerful method is to estimate the frequencies of half-sib (sharing a

746   single parent) and full-sib (sharing both parents) dyads, Q_{HS} and Q_{FS}, in a sample taken at random

747   from a single cohort of a population (Wang, 2009; Waples and Waples, 2011). Wang (2009)

748   derived a formula of N_e in terms of half- and full-sib frequencies, using both an inbreeding and a

749   drift approach,

750    $$\frac{1}{N_e} = \frac{1+3\alpha}{4}(Q_{HS} + 2Q_{FS}) - \frac{\alpha}{2}\left(\frac{1}{N_m} + \frac{1}{N_f}\right). \tag{27}$$

751   The equation has the parameter α as in eqn (5), so that the N_e for a population under non-random

752   mating (e.g. partial selfing) can be estimated. While Q_{HS} and Q_{FS} can be estimated from a sibship

753   assignment analysis of the multilocus genotypes (Wang and Santure, 2009; Jones and Wang, 1010),

754   α can be estimated from the same data with a F_{ST}-like approach (Wang, 2009). Alternatively, α can

755   be assumed to be zero for an outbred population when marker genotype frequencies do not deviate

756   significantly from those expected under Hardy-Weinberg equilibrium. The difficulty comes from

757   the estimation of the numbers of breeding males, N_m, and females, N_f, because sibship analysis

758   generally makes no distinctions between paternal and maternal sibships from autosomal marker

759   data, except in some specific situations (Wang, 2009). However, as detailed in Wang (2009), the

760   bias brought about by the last term of eqn (27), $\frac{\alpha}{2}\left(\frac{1}{N_m} + \frac{1}{N_f}\right)$, is usually negligible because α is

761   usually small and the estimate of $\frac{1}{N_m} + \frac{1}{N_f}$ by a sibship analysis is also not too far from its true

762   value.

763   There are several advantages of this sibship approach compared with other single-sample

764   approaches to N_e estimation. First, sibship can be inferred more accurately than other quantities

such as relatedness, which leads to more accurate estimates of N_e. Second, the approach applies to non-random mating populations, as an inbreeding coefficient α (eqn 5; equivalent to Wright's F_{IS}) can be calculated from the genotype data and incorporated into the N_e estimate. Similarly, the approach is also robust to population subdivision, as discussed by Wang (2009). Third, it applies to diploid species, haplodiploid species, dioecious as well as monoecious species with selfing. Fourth, a great advantage is that it provides not only an estimate of the summary parameter N_e, but also some information about the numbers of male and female parents and variance in family sizes through the sibship assignment analysis. This detailed information is especially valuable for conservation management, as a low N_e due to high variance in family size or a low number of parents would imply different management strategies. Simulations verified that the approach is much more accurate than the heterozygosity excess method, and is similar in accuracy to the temporal methods (Wang, 2009). However, it is unclear how its accuracy compares with that of the LD method. More work is needed to clarify this issue.

The above sibship frequency approach assumes a population with discrete generations. For a population with overlapping generations, the estimate provided by the sibship frequencies in a sample of single-cohort individuals is the effective number of breeders, N_b. This parameter summarizes the effects of variation in reproductive success between age classes, between sexes, and between individuals within an age-sex class on genetic drift in a single breeding season, instead of in a life time. It is less useful than N_e, and no population genetics equations are in terms of N_b. However, in the absence of an estimate of N_e, N_b provides some information about the risks of inbreeding and loss of genetic variation in conservation populations (e.g. Waples and Antao 2014; Whiteley *et al*. 2015). For the case of overlapping generations, Wang *et al*. (2010) proposed a parentage assignment method to estimate the N_e and generation interval from the sex, age, and multilocus genotype information of a single sample of individuals taken at random from a population. Essentially the method estimates the life table by parentage assignments, and both N_e and generation interval are then calculated from the life table. Simulations showed that the method yields unbiased and reasonably accurate estimates of N_e under realistic sampling and genotyping effort. Application of the method to empirical data yields sensible N_e estimates that are supported by other sources of information from the population (Kamath *et al*. 2015).

Multiple sources of drift/inbreeding information

The above approaches to N_e estimation use a single source of information, such as heterozygote excess, LD, temporal allele frequency changes, and sibship/parentage frequencies. Each piece of information reflects a facet of the stochastic process (genetic drift or inbreeding), and combining

800   multiple pieces of information may potentially allow for a better delineation of the process and thus

801   yield a more accurate estimate of N_e. Tallmon *et al*. (2008) proposed to use approximate Bayesian

802   computation (ABC) to estimate N_e from a sample of microsatellite genotypes. Their method,

803   implemented in a computer program ONESAMP, calculates and uses eight summary statistics

804   which are known to have functional relationships with N_e from population genetics theory or

805   simulations. These statistics include, among others, the number of alleles per locus, expected

806   heterozygosity, linkage disequilibrium, Wright's F_{IS}, the mean and variance of multilocus

807   homozygosity. In essence, the ABC approach simulates populations of different N_e and tries to find

808   the N_e value that yields the same or similar summary statistics to those calculated from the real data.

809   Tallmon *et al*. (2008) demonstrated this ABC approach by analysing an introduced increasing

810   population of ibex *Capra ibex*.

811   It is arguable that the ABC approach uses more information than other approaches. On one

812   hand, it uses multiple sources of information such as heterozygosity, number of alleles, and LD.

813   However, on the other hand, for each source of information, it uses a summary statistic rather than

814   the full information that is used by the probability methods (likelihood or Bayesian). Furthermore, it

815   is unclear how these different summary statistics should be optimally weighted, given that these

816   statistics are, apparently, highly correlated and may reflect the inbreeding and drift processes of

817   different time scales. For a population changing in size, these different summary statistics are

818   relevant for N_e in different time scales. For example, F_{IS} (like heterozygosity excess) is pertinent to

819   the parental N_e, LD implicates N_e in the past few or more generations (depending on the linkage of

820   the markers), while the number of alleles can be determined by the ancient N_e many generations (in

821   the order of N_e or $1/u$, whichever is smaller, where u is mutation rate) ago. So far an extensive

822   simulation study to compare the accuracy of ABC and other approaches is lacking, but is urgently

823   needed.

824

825   **Discussion**

826

827   Since the seminal work of Wright (1931, 1933), great progress has been made on the pivotal

828   population genetics parameter, N_e, in its concepts (e.g. inbreeding, variance, eigenvalue effective

829   sizes, etc.), its predictions for various species under different mating systems and population

830   structures, and its estimation methodologies using various marker information. In parallel, estimates

831   of N_e, from both demographic and genetic data, have been made for many populations in the past 30

832   years, thanks to the rapid developments in both molecular technologies and statistical and

833   computational methodologies.

834　　　　Much work has been done to predict the effect of selection on N_e. However, developing useful

835　　predictive models on the effect of selection acting on an inherited trait remains difficult. The reason

836　　is that the impact of linked genes propagates over a number of generations, resulting in a long-term

837　　effect that is difficult to combine in a simple equation with parameters referred to only one

838　　generation time. Coalescence theory runs into similar difficulties in predicting N_e, because the

839　　probabilities of coalescence for consecutive generations are not independent under selection on an

840　　inherited trait. In addition to the variation of N_e across generations, there is also variation over the

841　　genome. It is nowadays quite clear that there is a significant heterogeneity in levels and patterns

842　　of genetic variation across the genome caused by selection (Charlesworth, 2009; Gossmann *et al.*,

843　　2011), which complicates the inferences of N_e.

844　　　　Another important remaining problem about selection is the interrelationship between N_e and

845　　genetic variation. Most equations of N_e are linear functions of the census size N where genetic

846　　variation of the selected trait is an independent variable. However, genetic variation itself depends

847　　on N_e. Ignoring this fact is irrelevant for some purposes, but is troublesome in some models of

848　　closely linked loci. This reciprocal dependence is in the basis of the Hill-Robertson effect (Hill and

849　　Robertson, 1966) and Muller´s ratchet (Haigh, 1978), both being different aspects of the same issue,

850　　an additional reduction of genetic variance due to genetic drift induced by selection.

851　　　　　Demographic estimation of N_e can be made by application of the predictive equations

852　　reviewed here when information on census sizes, variances of progeny numbers, type of mating

853　　system, and other demographic data, are available. The lack of these data and the increasing

854　　availability of genetic markers make the estimation of N_e through genetic data to be, however, the

855　　leading procedure. Most factors affecting the populations in real situations imply a reduction of the

856　　effective size relative to the census size. In fact, the observed ratio N_e / N has been found to be

857　　about 10%-20% (Frankham, 1995; Palstra and Fraser, 2012) on average in meta-analyses across

858　　many species and populations. Overall, these figures are in agreement with theoretical expectations

859　　obtained from some of the predictive equations presented in this review when fluctuations in

860　　population size are considered (Vucetich *et al.*, 1997). However, this average N_e / N ratio may be an

861　　overestimate, as marine species are under-represented in these meta-analyses and can have

862　　extremely low N_e / N ratios.

863　　　　　Each N_e estimation method with genetic data is based on a certain population genetics model

864　　and has a number of assumptions. It is important to realize that, when these assumptions are

865　　violated, which is unfortunately the rule rather than the exception in the real world, an estimation

866　　method may yield invalid or biased estimates of N_e. For example, most methods reviewed herein

867　　assume an isolated random mating population with discrete generations. Yet in practice such

868　　populations are rare. The robustness of different methods has not been fully investigated.

869   Another important issue is the interpretation of the estimates obtained from a certain method.

870   First, what is the spatial scale relevant to a N_e estimate? Is it the effective size of the local

871   population from which samples are taken, or that of the metapopulation of which the sampled local

872   population is a part? Apparently, the local and metapopulation effective sizes are very different in

873   quantities and in applications. Local and metapopulation N_es signify the intensities of inbreeding

874   and genetic drift processes at the local and global levels, respectively. A small local N_e but a large

875   global N_e (i.e. many small interconnected populations) and a large local N_e but a small global N_e (i.e.

876   a few large interconnected populations) have different ecological, evolutionary, and conservation

877   genetics implications.

878   Second, what is the temporal scale relevant to a N_e estimate (Waples 2005)? Is it the N_e of the

879   sampled population, of the parental population, of an ancestral population or the average N_e over

880   several/many previous generations? The heterozygosity excess method and sibship methods

881   estimate the parental population N_e, temporal methods estimate the (harmonic) average N_e over the

882   generations in the sampling interval, while LD and ABC methods estimate the average N_e over an

883   unspecified number of previous generations (Wang, 2005). Of course the temporal scale becomes

884   irrelevant for a population with a constant unchanging demography. In practice, however, a natural

885   population never stays the same.

886   Third, does the estimate refer to inbreeding or variance effective size? The question is

887   irrelevant for the case of an isolated population with a constant demography or an incompletely

888   subdivided (i.e. with migration) population of constant size, as the inbreeding (N_{eI}) and variance

889   (N_{eV}) effective sizes are the same. Otherwise, however, N_{eI} and N_{eV} can be dramatically different. A

890   decreasing (increasing) population will always have a N_{eI} greater (smaller) than N_{eV}, because the

891   former and latter depend on the stochastic processes in the parental and the offspring generations,

892   respectively (Crow and Kimura, 1970; Caballero, 1994; Wang and Caballero, 1999). Although

893   usually unspecified in the original work describing the estimators, they estimate N_{eI}, N_{eV} or a

894   combination of both. The heterozygosity excess and sibship method estimate N_{eI}, while the

895   temporal methods estimate N_{eV}. It is unclear what the LD and ABC methods estimate, but they

896   likely estimate a combination of both N_{eI} and N_{eV}.

897   Fourth, are the estimators unbiased and accurate when their underlying assumptions are met

898   and are violated? Most estimators are not fully evaluated for their performances and statistical

899   behaviours by simulation studies, especially those that are computationally intensive. In measuring

900   the overall accuracy of an estimator, both precision and bias should be considered, and better

901   incorporated into a single measurement, such as mean squared errors. It is better to measure the

902   mean squared errors of $1/(2N_e)$ rather than N_e, because the latter can be infinitely large and, more

903   importantly, it is invariably $1/(2N_e)$ rather than N_e that enters a population genetics equation (Wang

904 and Whitlock, 2003). The dominating factor in determining accuracy is precision and bias when

905 marker information is scarce and ample, respectively.

906 More work is also needed in developing estimators that make fewer restrictive assumptions

907 and thus are more widely applicable to real populations. A common challenge to the current

908 estimators is population genetic structure, including age structure (i.e. overlapping generations) and

909 spatial structure (i.e. population subdivision with migration). Blindly applying an estimator

910 developed for a single isolated population with discrete generations under random mating to an age-

911 or space-structured population may yield unpredictable results.

912 Most current estimators have good performance for a population with a small N_e, partly

913 because the drift or inbreeding signal is strong. For a large population with N_e in the thousands or

914 more, drift and inbreeding in the population is weak and is thus difficult to pick up by the typical

915 sampling intensity (say, ~100 individuals, 10-20 microsatellites). With the rapid developments of

916 genotyping (sequencing) technology, an increasing number of studies use many genome-wide

917 markers to investigate population structure and demography. Use of thousands of SNPs may

918 compensate for the weak signal of stochasticity in a large population, and thus may allow for a good

919 estimate of large N_e. Future work is needed in evaluating current methods to estimate large N_e using

920 many markers. It can be reasonably assumed that, for an accurate estimate of N_e (say, a narrow 95%

921 confidence interval of roughly $[0.5N_e, 2N_e]$) by any efficient method, a sample size of individuals,

922 of the total number of effective alleles across loci, or of both in the order of N_e might be required.

923 A related issue with large N_e and many markers is the computational efficiency of likelihood

924 or Bayesian estimators. Efficient algorithms, such as that proposed recently for the likelihood

925 temporal estimator (Hui and Burt, 2015), are urgently needed to deal with big datasets. Another

926 option is to exploit the modern multicore and multi-cpu computers and to parallel computer

927 program codes using techniques such as MPI (Message Passing Interface) and OpenMP.

928 In some applications, the parameter, N_e, is all that is required in describing the current and

929 predicting the future genetic variation in a population, and the demographic details that determine

930 N_e are irrelevant. In other applications such as the conservation management of endangered species,

931 however, both N_e and the demographic details of the population are useful in designing the most

932 effective management to maintain the genetic diversity (Wang, 2009). When a population is

933 estimated to have a small N_e and thus is prone to the loss of genetic variation, we may ask what the

934 causes are. Is the small N_e caused by a small number of breeders, by a large variance in

935 reproductive success among breeders, by an unbalanced sex ratio, etc? Different causes imply

936 different optimal management strategies. The utility of future N_e estimators could be improved if

937 they provided joint estimates of N_e and important demographic quantities such as variance of

938 reproductive success.

29

939

Acknowledgements

945

References

947

Allendorf FW, Luikart GH, Aitken SN (2013). *Conservation and the Genetics of Populations*. John Wiley and Sons, Chichester, West Sussex, UK.

Anderson EC, Williamson EG, Thompson EA (2000). Monte Carlo evaluation of the likelihood for Ne from temporally spaced samples. *Genetics* **156**: 2109–2118.

Barbato M, Orozco-terWengel P, Tapio M, Bruford MW (2015). SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Frontiers in Genetics* **6**: 109.

Ballou JD, Lacy RC (1995). Identifying genetically important individuals for management of genetic diversity in pedigreed populations. In Ballou JD, Gilpin M, Foose TJ (eds.), *Population Management for Survival and Recovery*. Columbia University Press, New York.

Balloux F, Lehmann L (2003). Random mating with a finite number of matings. *Genetics* **165**: 2313–2315.

Balloux F, Lehmann L, De Meeus T (2003). The population genetics of clonal and partially clonal diploids. *Genetics* **164**: 1635–1644.

Beaumont MA (2003a). Conservation genetics. In *Handbook of statistical genetics* (ed. Balding DJ, Bishop M, Cannings C) 2nd edn., pp. 751–766. London: Wiley.

Beaumont MA (2003b). Estimation of population growth or decline in genetically monitored populations. *Genetics* **164**: 1139–1160.

Berthier P, Beaumont MA, Cornuet JM, Luikart G (2002). Likelihood-based estimation of the effective population size using temporal changes in allele frequencies: a genealogical approach. *Genetics* **160**: 741–751.

Bijma P, Van Arendonk JA, Woolliams JA. (2000). A general procedure for predicting rates of inbreeding in populations undergoing mass selection. *Genetics* **154**: 1865–1877.

Bijma P, Woolliams JA (1999). Prediction of genetic contributions and generation intervals in populations with overlapping generations under selection. *Genetics* **151**: 1197–1210.

973 Bijma P, Woolliams JA (2000). Prediction of rates of inbreeding in populations selected on best
974   linear unbiased prediction of breeding value. *Genetics* **156**: 361–373.

975 Bijma P, Van Arendonk JAM, Woolliams JA (2001). Predicting rates of inbreeding for livestock
976   improvement schemes. *J Anim Sci* **79**: 840–853.

977 Blouin MS, Parsons M, Lacaille V, Lotz S (1996). Use of microsatellite loci to classify individuals
978   by relatedness. *Mol Ecol* **5**: 393-401.

979 Bollback JP, York TL, Nielsen R (2008). Estimation of 2Nes from temporal allele frequency data.
980   *Genetics* **179**: 497–502.

981 Caballero A (1994). Developments in the prediction of effective population size. *Heredity* **73**: 657–
982   679.

983 Caballero A (1995). On the effective size of populations with separate sexes, with particular
984   reference to sex-linked genes. *Genetics* **139**: 1007–1011.

985 Caballero A, Hill WG (1992). Effective size of nonrandom mating populations. *Genetics* **130**: 909–
986   916.

987 Caballero A, Santiago E, Toro MA (1996a). Systems of mating to reduce inbreeding in selected
988   populations. *Anim Sci* **62**: 431–442.

989 Caballero A, Toro MA (2000). Interrelations between effective population size and other pedigree
990   tools for the management of conserved populations. *Genet Res* **75**: 331–343.

991 Caballero A, Toro MA (2002). Analysis of genetic diversity for the management of conserved
992   subdivided populations. *Conserv Genet* **3**: 289–299.

993 Caballero A, Wei M, Hill WG (1996b). Survival rates of mutant genes under artificial selection
994   using individual and family information. *J Genet* **75**: 63–80.

995 Charlesworth B (2001). The effect of life-history and mode of inheritance on neutral genetic
996   variability. *Genet Res* **77**: 153–166.

997 Charlesworth B (2009). Fundamental concepts in genetics: effective population
998   size and patterns of molecular evolution and variation. *Nat Rev Genet* **10**: 195–205.

999 Charlesworth B (2013). Background selection 20 years on: The Wilhelmine E. Key 2012
1000   invitational lecture. *J Hered* **104**: 161–171.

1001 Charlesworth B, Charlesworth D (2010). *Elements of Evolutionary Genetics*. Roberts & Co,
1002   Greenwood Village, Colorado, USA.

1003 Crow JF (1954). Breeding structure of populations. II. Effective population number. In *Statistics*
1004   *and Mathematics in Biology*, pp. 543-556. Iowa State COIL Press. Ames, Iowa.

1005 Crow JF, Morton NF (1955). Measurement of gene frequency drift in small populations. *Evolution*
1006   **9**: 202–214.

1007 Crow JF, Denniston C (1988). Inbreeding and variance effective population numbers. *Evolution* **42**:
1008   482–495.

1009 Crow JF, Kimura M (1970). *An Introduction to Population Genetics Theory*. Harper & Row, New
1010   York.

1011 Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR (2014). NeEstimator v2: re-
1012   implementation of software for the estimation of contemporary effective population size (Ne)
1013   from genetic data. *Mol Ecol Res* **14**: 209-214.

1014 Engen S, Lande R, Saether BE (2005). Effective size of a fluctuating age-structured population.
1015   *Genetics* **170**: 941–54.

1016 Evans BJ, Charlesworth B (2013). The effect of nonindependent mate pairing on the effective
1017   population size. *Genetics* 193: 545–556.

1018 Ewens WI (1979). *Mathematical Population Genetics*. Springer, Berlin.

1019 Falconer D, Mackay TFC (1996). *Introduction to Quantitative Genetics*. Longman, Harlow.

1020 Felsenstein J (1971). Inbreeding and variance effective numbers in populations with overlapping
1021   generations. *Genetics* **68**: 581–597.

1022 Fernández J, Toro MA, Caballero A (2003). Fixed contributions designs versus minimization of
1023   global coancestry to control inbreeding in small populations. *Genetics* **165**: 885–894.

1024 Fernández J, Toro MA, Caballero A (2004). Managing individuals' contributions to maximize the
1025   allelic diversity maintained in small, conserved populations. *Conserv Biol* **18**: 1–10.

1026 Fisher, RA (1965). *The Theory of Inbreeding*. Academic Press, New York, Second edition.

1027 Foll M, Shim H, Jensen JD (2015). WFABC: a Wright–Fisher ABC-based approach for inferring
1028   effective population sizes and selection coefficients from time-sampled data. *Mol Ecol Res* **15**:
1029   87–98.

1030 Frankham RD (1995). Effective population size / adult population size ratios in wildlife: a review.
1031   *Genet Res* **66**: 95-107.

1032 Frankham R, Ballou JD, Briscoe DA (2010). *Introduction to Conservation Genetics*. Cambridge
1033   University Press, Cambridge, UK.

1034 Ghai GL (1969). Structure of populations under mixed random and sib mating. *Theor Appl Genet*
1035   **39**: 179–182.

1036 Gilbert KJ, Whitlock MC. (2015). Evaluating methods for estimating local effective population size
1037   with and without migration. *Evolution* **69:** 2154-2166.

1038 Gillespie JH (2000). Genetic drift in an infinite population: the pseudo hitchhiking model. *Genetics*
1039   **155**: 909–919.

1040 Gossmann TI, Woolfit M, Eyre-Walker A (2011). Quantifying the variation in the effective
1041   population size within a genome. *Genetics* **189**: 1389–1402.

1042 Gowe RS, Robertson A, Latter BDH (1959). Environment and poultry breeding problems. 5. The
1043   design of poultry control strains. *Poul Sci* **38**: 462–471.

1044 Grimm A, Gruber B, Hoehn M, Enders K, Henle K (2016). A model-derived short-term estimation
1045   method of effective size for small populations with overlapping generations. *Methods in*
1046   *Ecology and Evolution* (in press).

1047 Haigh J (1978). The accumulation of deleterious genes in a population–Muller's Ratchet. *Theor*
1048   *Popul Biol* **14**: 251–267.

1049 Haldane JBS (1924). A mathematical theory of natural and artificial selection. H. The influence of
1050   partial self-fertilisation, inbreeding, assortative mating, and selective fertilisation on the
1051   composition of Mendelian populations, and on natural selection. *Proc. Camb. Phil. Soc., (Biol.*
1052   *Sci.), (later Biol. Rev.)* **1**: 158–163.

1053 Haldane JBS (1939). The equilibrium between mutation and random extinction. *Ann Eugen* **9**: 400–
1054   405.

1055 Hare MP, Nunney L, Schwartz MK, *et al*. (2011). Understanding and estimating effective
1056   population size for practical application in marine species management. *Conserv Biol* **25:** 438-
1057   449.

1058 Hayes BJ, Visscher PM, McPartlan HC, Goddard ME (2003). Novel multilocus measure of linkage
1059   disequilibrium to estimate past effective population size. *Genome Res* **13**: 635–643.

1060 Hedrick P, Jain S, Holden L (1978). Multilocus systems in evolution. *Evol Biol* **11**: 101-184.

1061 Hill WG (1972). Effective size of populations with overlapping generations. *Theor Popul Biol* **3**:
1062   278–289.

1063 Hill WG (1979). A note on effective population size with overlapping generations. *Genetics* **92**:
1064   317–322.

1065 Hill WG (1981). Estimation of effective population size from data on linkage disequilibrium. *Genet*
1066   *Res* **38**: 209–216.

1067 Hill WG, Robertson A (1966). The effects of linkage on limits to artificial selection. *Genet Res* **8**:
1068   269–294.

1069 Hudson RR, Kaplan NL (1995). Deleterious background selection with recombination. *Genetics*
1070   **141**: 1605–1617.

1071 Hui TYJ, Burt A (2015). Estimating effective population size from temporally spaced samples with
1072   a novel, efficient maximum-likelihood algorithm. *Genetics* **200**: 285-293.

1073 Jones OR, Wang J (2010). COLONY: a program for parentage and sibship inference from
1074   multilocus genotype data. *Mol Ecol Res* **10**: 551-555.

1075   Jorde PE, Ryman N (1995). Temporal allele frequency change and estimation of effective size in
1076   populations with overlapping generations. *Genetics* **139**: 1077–1090.

1077   Kamath PL, Haroldson MA, Luikart G, Paetkau D, Whitman C, Manen FT (2015). Multiple
1078   estimates of effective population size for monitoring a long-lived vertebrate: an application to
1079   Yellowstone grizzly bears. *Mol. Ecol.* **24**, 5507-5521.

1080   Kimura M, Crow JF (1963a). The measurement of effective population number. *Evolution* **17**: 279–
1081   288.

1082   Kimura M, Crow JF (1963b). On the maximum avoidance of inbreeding. *Genet Res* **4**: 399–415.

1083   Krimbas CB, Tsakas S (1971). The genetics of *Dacus oleae* V. Changes of esterase polymorphism
1084   in a natural population following insecticide control: selection or drift? *Evolution* **25**: 454–460.

1085   Lacy R (1995). Clarification of genetic terms and their use in the management of captive
1086   populations. *Zoo Biol* **14**: 565–578.

1087   Laporte V, Charlesworth B (2002). Effective population size and population subdivision in
1088   demographically structured populations. *Genetics* **162**: 501–519.

1089   Laporte V, Cuguen J, Couvet D (2000). Effective population sizes for cytoplasmic and nuclear
1090   genes in a gynodioecious species: the role of the sex determination system. *Genetics* **154**: 447–
1091   458.

1092   Laval G, SanCristobal M, Chevalet C (2003). Maximum-likehhood and Markov chain Monte Carlo
1093   approaches to estimate inbreeding and effective size from allele frequency changes. *Genetics*
1094   **164**: 1189–1204.

1095   Li CC (1976). *First Course in Population Genetics*. Boxwood, Pacific Grove, CA.

1096   Luikart G, Cornuet JM (1999). Estimating the effective number of breeders from heterozygote
1097   excess in progeny. *Genetics* **151**: 1211–1216.

1098   Luikart G, Ryman N, Tallmon DA, Schwartz MK, Allendorf FW (2010). Estimation of census and
1099   effective population sizes: the increasing usefulness of DNA-based approaches. *Conserv*
1100   *Genet* **11**: 355-373.

1101   Mathieson I, McVean G (2013). Estimating selection coefficients in spatially structured populations
1102   from time series data of allele frequencies. *Genetics* **193**: 973–984.

1103   Mezzavilla M, Ghirotto S (2015). Neon: An R package to estimate human effective population size
1104   and divergence time from patterns of linkage disequilibrium between SNPs. *J Comput Sci Syst*
1105   *Biol* **8**: O37-O44.

1106   Meuwissen THE (2007). Operation of conservation schemes. In: Oldenbroek, K. (Ed.), Utilisation
1107   and Conservation of Farm Animal Genetic Resources. Wageningen Academic Publishers,
1108   Wageningen, The Netherlands, pp. 167–193.

1109 Nagylaki T (1992). *Theoretical Population Genetics*. Springer, Berlin.

1110 Nei M, Murata M (1966). Effective population size when fertility is inherited. *Genet Res* **8**: 257–
1111   260.

1112 Nei M, Tajima F (1981). Genetic drift and estimation of effective population size. *Genetics* **98**:
1113   625–640.

1114 Nicolaisen LE, Desai MM (2012). Distortions in genealogies due to purifying selection. *Mol Biol*
1115   *Evol* **29**: 3589–3600.

1116 Nicolaisen LE, Desai MM (2013). Distortions in genealogies due to purifying selection and
1117   recombination. *Genetics* **195**: 221–230.

1118 Nomura T (1996). Effective size of selected populations with overlapping generations. *J Anim*
1119   *Breed Genet* **113**: 1–16.

1120 Nomura T (1997a). On the effective size of nonrandom mating populations under selection. *Jpn J*
1121   *Biometrics* **18**: 1–11.

1122 Nomura T (1997b). Effective population size for a sex-linked locus in populations under selection.
1123   *Math Biosci* **142**: 79–89.

1124 Nomura T (1997c). Prediction of effective population size in open nucleus breeding systems. *J*
1125   *Anim Breed Genet* **114**: 333–347.

1126 Nomura T (1998a). Effective population size in supportive breeding. *Conserv Biol* **13**: 670–672.

1127 Nomura T (1998b). Inbreeding and effective size of populations under selection. Proceedings of the
1128   MAFF International Workshop on Genetic Resources Animal Genetic Resources II, pp. 5–24,
1129   National Institute of Agrobiological Resources, Tsukuba.

1130 Nomura T (1999a). Inbreeding coefficient and effective size of finite populations with a specified
1131   proportion of sib-mating. *Jpn J Biometrics* **83**: 485–489.

1132 Nomura T (1999b). On the methods for predicting the effective size of populations under selection.
1133   *Heredity* **83**: 485–489.

1134 Nomura T (2000). Effective population size under marker assisted selection. *Jap J Biometrics* **21**:
1135   1–12.

1136 Nomura T (2002a). Effective size of populations with unequal sex ratio and variation in mating
1137   success. *J Anim Breed Genet* **118**: 297–310.

1138 Nomura T (2002b). Effective size of populations with heritable variation in fitness. *Heredity* **89**:
1139   413–416.

1140 Nomura T (2005a). Developments in prediction theories of the effective size of populations under
1141   selection. *Anim Sci J* **76**: 87–96.

1142 Nomura T (2005b). Effective population size under random mating with a finite number of matings.
1143   *Genetics* **171**: 1441–1442.

Nomura T (2008). Estimation of effective number of breeders from molecular coancestry of single cohort sample. *Evol Appl* **1**: 462–474.

Nomura T, Mukai F, Yamamoto A (1999). Prediction of response and inbreeding under selection based on best linear unbiased prediction in closed broiler lines. *Animal Science Journal* **70**: 273–281.

Nomura T, Takahashi J (2012). Effective population size in eusocial Hymenoptera with worker-produced males. *Heredity* **109**: 261–268.

Nomura T, Yonezawa K (1996). A comparison of four systems of group mating for avaiding inbreeding. *Genet Sel Evol* **28**: 141–159.

Nordborg M, Krone SM (2002). Separation of time scales and convergence to the coalescent in structured populations. In *Modern Developments in Theoretical Population Genetics: The Legacy of Gustave Malécot*. pp 194–232. Edited by Slatkin M, Veuille M. Oxford University Press, Oxford.

Nordborg M, Charlesworth B, Charlesworth D (1996). The effect of recombination on background selection. *Genet Res* **67**: 159–174.

Nunney L (1993). The influence of mating system and overlapping generations on effective population size. *Evolution* 47: 1329–1341.

Orive M (1993). Effective population size in organisms with complex life-histories, *Theor Pop Biol* **44**: 316–340.

Palstra FP, Ruzzante DE (2008). Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Mol Ecol* **17**: 3428–3447.

Palstra FP, Fraser DJ (2012). Effective/census population size ratio estimation: a compendium and appraisal. *Ecol Evol* **2**: 2357-2365.

Pudovkin AI, Zaykin DV, Hedgecock D (1996). On the potential for estimating the effective number of breeders from heterozygote-excess in progeny. *Genetics* **144**: 383–387.

Pollak E (1983). A new method for estimating the effective population size from allele frequency changes. *Genetics* **104**: 531–548.

Pollak E (1987). On the theory of partially inbreeding finite populations. I. Partial selling. *Genetics* **117**: 353–360.

Pollak E (2002). Eigenvalue effective population numbers for populations that vary cyclically in size. *Math Biosci* **177-178**: 11–24.

Robertson A (1961). Inbreeding in artificial selection programmes. *Genet Res* **2**: 189–194.

Robertson A (1964). The effect of nonrandom mating within inbred lines on the rate of inbreeding. *Genet Res* **5**: 164–167.

1179   Robertson A (1965). The interpretation of genotypic ratios in domestic animal populations. *Anim*
1180   *Prod* **7**: 319–324.

1181   Rönnegård L, Woolliams JA (2003). Predicted rates of inbreeding with additive maternal effects.
1182   *Genet Res* **82**: 67–77.

1183   Ryman N, Laikre L (1991). Effects of supportive breeding on the genetically effective population
1184   size. *Conserv Biol* **5**: 325–329.

1185   Ryman N, Jorde PE, Laikre L (1995). Supportive breeding and variance effective population size.
1186   *Conserv Biol* **9**: 1619–1628.

1187   Ryman N, Jorde PE, Laikre L (1999). Supportive breeding and inbreeding effective number: Reply
1188   to Nomura. *Conserv Biol* **13**: 673–676.

1189   Sánchez L, Bijma P, Woolliams JA (2003). Minimizing inbreeding by managing genetic
1190   contributions across generations. *Genetics* **164**: 1589–1595.

1191   Santiago E, Caballero A (1995). Effective size of populations under selection. *Genetics* **139**: 1013–
1192   1030.

1193   Santiago E, Caballero A (1998). Effective size and polymorphism of linked neutral loci in
1194   populations under directional selection. *Genetics* **149**: 2105–2117.

1195   Santiago E, Caballero A (2001). Application of reproductive technologies to the conservation of
1196   genetic resources. *Conserv Biol* **14**: 1831–1836.

1197   Saura M, Tenesa A, Woolliams JA, Fernández A, Villanueva B. (2015). Evaluation of the linkage-
1198   disequilibrium method for the estimation of effective population size when generations
1199   overlap: an empirical case. *BMC Genomics* **16:** 922.

1200   Schwartz MK, Tallmon DA, Luikart G (1999). Using genetics to estimate the size of wild
1201   populations: many methods, much potential, uncertain utility. *Anim Conserv* **2**: 321–323.

1202   Schwartz MK, Luikart G, Waples RS (2007). Genetic monitoring as a promising tool for
1203   conservation and management. *TREE* **22**: 25-33.

1204   Sved JA (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite
1205   populations. *Theor Pop Biol* **2**: 125-141.

1206   Tallmon DA, Luikart G, Beaumont MA (2004). Comparative evaluation of a new effective
1207   population size estimator based on approximate Bayesian computation. *Genetics* **167**: 977–
1208   988.

1209   Tallmon DA, Koyuk A, Luikart G, Beaumont MA (2008). COMPUTER PROGRAMS: onesamp: a
1210   program to estimate effective population size using approximate Bayesian computation. *Mol*
1211   *Ecol Res* **8**: 299-301.

Tallmon DA, Gregovich D, Waples RS, Baker CS, Jackson J, *et al*. (2010) When are genetic methods useful for estimating contemporary abundance and detecting population trends? *Mol. Ecol. Res.* **10**: 684-692.

Theodorou K, Couvet D (2010). Genetic management of captive populations: the advantages of circular mating. *Conserv Genet* **11**: 2289–2297.

Toro MA, Pérez-Enciso M (1990). Optimization of selection response under restricted inbreeding. *Genet Sel Evol* **22**: 93–107.

Vencovsky R, Chaves LJ, Crossa J (2012). Variance population size for dioecious species. *Crop Sci* **52**: 79–90.

Vucetich JA, Waite TA, Nunney L (1997). Fluctuating population size and the ratio of effective size to census population size. *Evolution* **51**: 2017–3021.

Wakeley J (2008). *Coalescent Theory. An Introduction.* Ben Roberts, Greenwood Village, Colorado, USA.

Wakeley J, Sargsyan O (2009). Extensions of the coalescent effective population size. *Genetics* **181**: 341–345.

Wang J (1996a). Deviation from Hardy-Weinberg proportions in finite populations. *Genet Res* **68**: 249–257.

Wang J (1996b). Inbreeding and variance effective sizes for non-random mating populations. *Evolution* **50**: 1786–1794.

Wang J (1996c). Inbreeding coefficient and effective size for an X-linked locus in nonrandom mating populations. *Heredity* **76**: 569–577.

Wang J (1997a). Effective size and F-statistics of subdivided populations. I. Monoecious species with partial selfing. *Genetics* **146**: 1453–1463.

Wang J (1997b). Effective size and F-statistics of subdivided populations. II. Dioecious species. *Genetics* **146**: 1465–1474.

Wang J (1997c). More efficient breeding systems for controlling inbreeding and effective size in animal populations. *Heredity* **79**: 591–599.

Wang J (1997d). Effect of excluding sib matings on inbreeding coefficient and effective size of finite diploid populations. *Biometrics* **53**: 1354–1365.

Wang J (1998). Effective size of populations under selection: some extensions. *Proc 6th World Cong Genet Appl Lives Prod* **26**: 13–16.

Wang J (2001a). Optimal marker-assisted selection to increase the effective size of small populations. *Genetics* **157**: 867–874.

Wang J (2001b). A pseudo-likelihood method for estimating effective population size from temporally spaced samples. *Genet Res* **78**: 243–257.

1247 Wang J (2004). Sibship reconstruction from genetic data with typing errors. *Genetics* **166**: 1963-
1248   1979.

1249 Wang J (2005). Estimation of effective population sizes from data on genetic markers. *Phil Trans R*
1250   *Soc B* **360**: 1395–1409.

1251 Wang J (2009). A new method for estimating effective population sizes from a single sample of
1252   multilocus genotypes. *Mol Ecol* **18**: 2148–2164.

1253 Wang J (2014). Marker-based estimates of relatedness and inbreeding coefficients: an assessment of
1254   current methods. *J Evol Biol* **27**: 518-530.

1255 Wang J, Caballero A (1999). Developments in predicting the effective size of subdivided
1256   populations. *Heredity* **82**: 212–226.

1257 Wang J, Hill WG (2000). Marker assisted selection to increase effective population size by reducing
1258   Mendelian segregation variance. *Genetics* **154**: 475–489.

1259 Wang J, Ryman N (2001). Genetic effects of multiple generations of supportive breeding. *Conserv*
1260   *Biol* **15**: 1619-1631.

1261 Wang J, Santure AW (2009). Parentage and sibship inference from multilocus genotype data under
1262   polygamy. *Genetics* **181***:* 1579-1594.

1263 Wang J, Whitlock MC (2003). Estimating effective population size and migration rates from
1264   genetic samples over space and time. *Genetics* **163**: 429-446.

1265 Wang J, Brekke P, Huchard E, Knapp LA, Cowlishaw G (2010). Estimation of parameters of
1266   inbreeding and genetic drift in populations with overlapping generations. *Evolution* **64**: 1704–
1267   1718.

1268 Wang Y, Pollak E (2002). The effective number of population that varies cyclically in size. II.
1269   Overlapping generations. *Math Biosci* **179**: 161–181.

1270 Waples RS (1989). A generalized approach for estimating effective population size from temporal
1271   changes in allele frequency. *Genetics* **121**: 379–391.

1272 Waples RS (2005) Genetic estimates of contemporary effective population size: to what time
1273   periods do the estimates apply? *Mol Ecol* 14: 3335-3352.

1274 Waples RS (2006). A bias correction for estimates of effective population size based on linkage
1275   disequilibrium at unlinked gene loci. *Conserv Genet* **7**: 167–184.

1276 Waples RS, Do C (2008). LdNe: a program for estimating effective population size from data on
1277   linkage disequilibrium. *Mol Ecol Res* 8: 753–756.

1278 Waples RS, Do C (2010). Linkage disequilibrium estimates of contemporary Ne using SNPs and
1279   highly polymorphic molecular markers: an evaluation of precision and bias. *Evol Appl* **3**: 244-
1280   262.

Waples RS, England PR (2011). Estimating contemporary effective population size on the basis of linkage disequilibrium in the face of migration. *Genetics* **189**: 633–644.

Waples RS, Antao T (2014). Intermittent breeding and constraints on litter size: consequences for effective population size per generation (N_e) and per reproductive cycle (N_b). *Evolution* **68**: 1722-1734.

Waples RS, Waples RK (2011). Inbreeding effective population size and parentage analysis without parents. *Mol Ecol Res* **11**: 162–171.

Waples RS, Yokota M (2007). Temporal estimates of effective population size in species with overlapping generations. *Genetics* **175**: 219–233.

Waples RS, Antao T, Luikart G (2014). Effects of overlapping generations on linkage disequilibrium estimates of effective population size. *Genetics* **197**: 769–780.

Weir BS, Hill WG (1980). Effect of mating structure on variation in linkage disequilibrium. *Genetics* **95**: 477-488.

Whiteley AR, Coombs JA, Cembrola M, O'Donnell MJ, Hudy M, Nislow KH, Letcher BH (2015). Effective number of breeders provides a link between interannual variation in stream flow and individual reproductive contribution in a stream salmonid. *Mol. Ecol.* **24**: 3585-3602.

Whitlock M, Barton N (1997). The effective size of a subdivided population. *Genetics* **146**: 427–441.

Wiehe THE, Stephan W (1993). Analysis of a genetic hitch-hiking model, and its application to DNA polymorphism data from *Drosophila melanogaster*. *Mol Biol Evol* **10**: 842–854.

Williamson EG, Slatkin M (1999). Using maximum likelihood to estimate population size from temporal changes in allele frequencies. *Genetics* **152**: 755–761.

Woolliams JA, Bijma P (2000). Predicting rates of inbreeding in populations undergoing selection. *Genetics* **154**: 1851–1864.

Woolliams JA, Thompson R (1994). A theory of genetic contributions. *Proc 5th World Cong Genet Appl Lives Prod* **19**: 127–134.

Woolliams JA, Wray NR, Thompson R (1993). Prediction of long-term contributions and inbreeding in populations undergoing mass selection. *Genet Res* **62**: 231–242.

Wray NR, Thompson R (1990). Predictions of rates of inbreeding in selected populations. *Genet Res* **55**: 41–54.

Wray NR, Woolliams JA, Thompson R (1990). Methods for predicting rates of inbreeding in selected populations. *Theor Appl Genet* **80**: 503–512.

Wray NR, Woolliams JA, Thompson R (1994). Predicting of rates of inbreeding in populations undergoing index selection. *Theor Appl Genet* **87**: 878–892.

Wright S (1921). Systems of mating. *Genetics* **6**: 111–178.

1316   Wright S (1931). Evolution in Mendelian populations. *Genetics* **16**: 97–159.

1317   Wright S (1933). Inbreeding and homozygosis. *Proc Natl Acad Sci USA* **19**: 411–420.

1318   Wright S (1938). Size of population and breeding structure in relation to evolution. *Science* **87**:
1319   430–431.

1320   Wright S (1939). *Statistical genetics in relation to evolution. Exposés de Biométrie et de Statistique*
1321   *Biologique*. Herman & Cie, Paris, France.

1322   Wright S (1943). Isolation by distance. Genetics 28: 114-138.

1323   Wright S (1969). *Evolution and the Genetics of Populations, vol. 2, The Theory of Gene*
1324   *Frequencies*. Chicago: University of Chicago Press.

1325   Yonezawa K (1997). Effective population size of plant species propagating with mixed sexual and
1326   asexual reproduction system. *Genet Res* **70**: 251–258.

1327   Yonezawa K, Kinoshita E, Watano Y, Zentoh H (2000). Formulation and estimation of the effective
1328   size of stage-structured populations in *Fritillaria camtschatcensis*, a perennial herb with a
1329   complex life history. *Genetics* **54**: 2007–2013.

1330   Yonezawa K, Ishii T, Nagamine T (2004). The effective size of mixed sexually and asexually
1331   reproducing populations. *Genetics* **166**: 1529–1539.

1332   Zhdanova OL, Pudovkin AI (2008). Nb_HetEx: a program to estimate the effective number of
1333   breeders. *J Hered* **99**: 694-695.