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*

Left running head: J Wang, E Santiago, A Caballero

Right running head: Predicting and Estimating N_e

Key words: effective population size, inbreeding, genetic drift, markers, conservation

Corresponding author:

Jinliang Wang
Institute of Zoology
Regent’s Park
London NW1 4RY
United Kingdom
Tel: 0044 20 74496620
Fax: 0044 20 75862870
Email: jinliang.wang@ioz.ac.uk

ABSTRACT

Effective population size (N_e) is a key parameter in population genetics. It has important applications in evolutionary biology, conservation genetics, and plant and animal breeding, because it measures the rates of genetic drift and inbreeding and affects the efficacy of systematic evolutionary forces such as mutation, selection and migration. We review the developments in predictive equations and estimation methodologies of effective size. In the prediction part, we focus on the equations for populations with different modes of reproduction, for populations under selection for unlinked or linked loci, and for the specific applications to conservation genetics. In the estimation part, we focus on methods developed for estimating the current or recent effective size from molecular marker or sequence data. We discuss some underdeveloped areas in predicting and estimating N_e for future research.

Introduction

The concept of effective population size, introduced by Sewall Wright (1931, 1933), is central to plant and animal breeding (Falconer and Mackay, 1996), conservation genetics (Frankham *et al*., 2010; Allendorf *et al*., 2013), and molecular variation and evolution (Charlesworth and Charlesworth, 2010), as it quantifies the magnitude of genetic drift and inbreeding in real-world populations. A substantial number of extensions to the basic theory and predictions were made since the seminal work of Wright, with main early developments by James Crow and Motoo Kimura (Kimura and Crow, 1963a; Crow and Kimura, 1970) and later by a list of contributors. Several review papers (Crow and Denniston, 1988; Caballero, 1994; Wang and Caballero, 1999; Nomura, 2005a), and population genetics books (Fisher, 1965; Wright, 1969; Ewens, 1979; Nagylaki, 1992) have summarised the existing theory in predicting the effective size of a population at different spatial and time scales under various inheritance modes and demographies. Comparatively, methodological developments (reviewed by Schwartz *et al*., 1999; Beaumont, 2003a; Wang, 2005; Palstra and Ruzzante, 2008; Luikart *et al*., 2010; Gilbert and Whitlock, 2015) in estimating the effective size of natural populations from genetic data lag behind, but are accelerating in the past decade thanks to the rapid developments of molecular biology.

The classical developments of effective population size theory are based on the rate of change in gene frequency variance (genetic drift) or the rate of inbreeding. The effective population size is defined in reference to the Wright-Fisher idealised population, i.e. a hypothetical population with very simplifying characteristics where genetic drift is the only factor in operation, and the dynamics of allelic and genotypic frequencies across generations merely depend on the population census (N) size. The effective size of a real population is then defined as the size of an idealised population which would give rise to the rate of inbreeding and the rate of change in variance of gene frequencies actually observed in the population under consideration, which correspond to the so-called *inbreeding and variance effective sizes*, respectively (Crow and Kimura, 1970).

Predictions of the effective population size can also be obtained from the largest nonunit eigenvalue of the transition matrix of a Markov Chain which describes the dynamics of allele frequencies. Such derived effective size is called *eigenvalue effective size* (see Ewens, 1979, p. 104-112), which is equivalent to the *random extinction effective size* (Crow, 1954; see also Haldane, 1939). The transition matrix can be written for many genetic models, and is particularly useful for complex scenarios such as populations varying in size, having age structures or being subject to demographic changes (e.g., Wang and Pollak, 2002; Charlesworth, 2001; Pollak, 2002; Engen *et al*., 2005). A less often used approach is that for the *mutation effective size*, defined by the probability of identity in state of genes rather than identity by descent under an infinite allele model of mutations with a defined mutation rate (Whitlock and Barton, 1997).

Later developments based on coalescence theory (Wakeley, 2008) have also proved to be useful in the prediction of effective population size, particularly in the evolutionary context for predicting genetic variability at the molecular level (Charlesworth, 2009; Nicolaisen and Desai, 2012, 2013). The coalescence theory states that the chance of coalescence of any two random gene copies in one generation time is $1/2N$, which is the same as the rate of increase in identity by descent occurred from one generation to the next one. Thus, the probability of coalescence t generations ago is $[1-(1/2N)]^{t-1}(1/2N)$. Therefore, the average time of coalescence of two 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 population size* refers to the expected time of coalescence T, in generations, of gene copies such that $T = 2N_e$ (Nordborg and Krone, 2002; Wakeley and Sargsyan, 2009).

In this paper, we present a general overview of the main developments for predicting the effective population size (N_e). The review does not attempt to be exhaustive, and some of the material mentioned in previous reviews will not be repeated. We mainly focus on populations with different modes of reproduction, populations under selection, and populations under genetic management in captive breeding conservation programmes, complementing previous reviews and adding material not covered or only partially covered by them. We also review the developments in estimating contemporary effective sizes from genetic marker data, focusing on the estimation principles and ignoring the technical details that were covered in the original papers. The underlying assumptions, application scopes, robustness and accuracies of different estimation methods are discussed and compared.

Prediction of the effective population size

In this section, we will summarize the main predictive equations for the asymptotic effective population size reached after a number of generations in a regular breeding system. In this case, all of the above approaches generally lead to the same predictive equations of N_e except for a few particular scenarios. For example, in a regular breeding system for an undivided population the asymptotic inbreeding and variance effective sizes converge. Only in situations such as when the population is subdivided permanently in independent sublines with completely independent pedigrees (Wang, 1997a, b), or when the population is decreasing or increasing in size, these types of N_e will differ permanently. In fact, many of the equations shown below have been derived by two or more of the above approaches, and we will only mention some of these. For clarity and a better understanding of the main principles, several simplifying assumptions will also be made in this prediction section. Unless otherwise stated, we will assume that populations do not change size through time and are large enough so that second order terms of $1/N_e$ can be safely ignored. These terms are generally of little relevance, but make the derivations and the N_e equations rather cumbersome. Finally, a single undivided population with discrete generations under a regular breeding scheme will be generally assumed unless otherwise indicated, so that prediction equations refer to asymptotic (Caballero, 1994) effective population sizes.

Populations with different modes of reproduction

As a starting point, we consider the simple equation derived by Wright (1938) which takes account of the variance of the contributions from parents to progeny (S_k^2) in a population of constant size N,

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

This expression also assumes a population either containing only hermaphrodites or comprising equal numbers of males and females, diploid autosomal inheritance, and random mating (including selfing for hermaphrodites). In eqn (1), the term S_k^2 accounts for the genetic drift caused by the variable contributions among parents, whereas the first term “2” in the denominator accounts for the genetic drift caused by the Mendelian segregation of heterozygotes (i.e. the drift in allele frequency arising from the fact that the progeny from a heterozygote can alternatively receive one or the other allele). It can also be seen as the variance in contributions between paternal and maternal genes at a locus within an individual or part of the variance in contribution between grandparents (the term δ^2 in eqn 2 of Wang and Hill [2000]).

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

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

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

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

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 the covariance of the numbers of male and female offspring from parents of sex *x*. Eqn (2) is the same as that derived by Hill (1979), although it is expressed in a different form. It reduces to the classical equation of Wright (1933, 1939) for a Poisson distribution of progeny number (i.e. $S_{xy}^2 = N_y / N_x, S_{xm,xf} = 0$ for sex *x*, *y* = *m*, *f*),

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

which shows that unequal numbers of males and females in a population introduce a systematic variance in contribution between male and female parents and thus a reduction in effective size.

Predictive formulae for the effective size of X-linked genes were originally given by Wright (1933) and later extended by other authors (see Caballero, 1995). Later developments have also been made for Y-linked and maternally transmitted genes (Charlesworth, 2001; Laporte and Charlesworth, 2002; Evans and Charlesworth, 2013).

The generalisation of eqn (1) to the case of a partially selfed population (in which there is partial selfing with proportion β, random mating otherwise), is

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

(Crow and Morton, 1955), where

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

(Haldane, 1924) quantifies the deviation from Hardy-Weinberg equilibrium or the correlation of genes within individuals relative to the genes taken at random from the population (Wright, 1969). The value of α in a large random mating population is approximately zero (slightly negative when second order terms are considered; see eqn (23) below and Wang, 1996a). For the case of biparental inbreeding such as partial full-sib mating in dioecious species, the expression is the same as (5) except that the term $(1 + \alpha)$ should be replaced by $(1 + 3\alpha)$ (Caballero and Hill, 1992). The equilibrium value of α for biparental inbreeding is also different from (6) (e.g. Ghai, 1969).

The generalisation of eqn (5) for different numbers of males and females was made by Wang (1996b), adding some covariance terms not considered by Caballero (1994; eqn 17). Wang's (1996b) equation also allows for different numbers of males and females varying over generations.

An excess of inbred matings ($\alpha > 0$ in eqn [5]) has the effect of increasing the correlation of genes within individuals and decreasing the frequency of heterozygotes by a factor of α. It results in a decrease in the genetic drift due to a decrease in Mendelian segregation variance, and an increase in the genetic drift due to an increase in the variance of contributions among individuals. Compared with random mating ($\alpha = 0$), whether an excess ($\alpha > 0$) or a deficit ($\alpha < 0$) of inbred matings may increase or decrease N_e depends on the variance in family size, S_k^2. For the case of partial selfing (eqn 5), inbreeding ($\alpha > 0$) increases N_e when $S_k^2 < 2$ and decreases N_e when $S_k^2 > 2$. At exactly $S_k^2 = 2$, selfing has no effect on N_e.

Predictions of the effective size for X-linked genes in non-random mating populations were given by Wang (1996c). Nomura (2002a, 2005b) also provided predictions of the effective size for a variety of mating systems in animals (see also Balloux and Lehmann [2003]). For example, for harem polygamy, where successful males generally mate with most or all of the females in their harem, and the females generally mate with only one male, the effective size, for a Poisson distribution of progeny number, is better approximated by

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

rather than by eqn (4), showing the larger impact of male number for this type of mating system. Other predictions of N_e for different mating systems and overlapping generations have been provided by Nunney (1993).

Expression (5) can also be obtained following the concept of long-term contributions from ancestors to descendants developed by Wray and Thompson (1990) in the context of populations under selection. As suggested by Woolliams and Thompson (1994) and shown by Caballero and Toro (2000, 2002), the expressions can be approximated by

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

where V_∞ is the variance of long-term contributions from ancestors to descendants. For random mating ($\alpha = 0$), $V_\infty = 1$ and $N_e = N$, as expected.

For a proportion β of partial selfing, $V_\infty \approx S_k^2/[2(1-\beta)]$ if α is not too close to one (Caballero and Toro, 2000), which, when substituted into (8), gives eqn (5). When the numbers of selfed and nonselfed progeny are independently Poisson distributed, $S_k^2 \approx 2 + 2\beta$ (see Nomura, 1999a for a more precise prediction), and both eqn (5) and (8) reduce to

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

(Li, 1976, p. 562; Pollak, 1987; Caballero and Hill, 1992; Nordborg and Krone, 2002). For a population with Poisson distribution of family size and complete selfing, (9) reduces to $N_e = N/2$. Equation (8) can also be applied to the scenario of partial full-sib mating with the appropriate approximation for V_∞ (Caballero and Toro, 2000).

Predictions of the effective size for populations with mixed sexual and asexual reproduction systems and discrete and overlapping generations have been developed by Yonezawa (1997). Assuming a monoecious plant species carrying out asexual propagation with a proportion δ in a population of constant size N (i.e. an average progeny number of one and two for asexual propagation and sexual reproduction, respectively), eqn (5) of Yonezawa (1997) can be rearranged to

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

where S_c^2 is the variance of the number of asexually produced progeny among plants. If there is no asexual reproduction (δ = 0), eqn (10) reduces to (5), as it should. If all reproduction is asexual (δ = 1), (10) reduces to

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

Interestingly, if the number of asexually produced progeny is Poisson distributed ($S_c^2 = 1$), the expression is the same as for a sexually reproducing partially selfed population where the numbers of selfed and nonselfed progeny are independently Poisson distributed, i.e. eqn (9). If all individuals are homozygotes ($\alpha = 1$), $N_e = N/2$, the same as for a fully selfed population.

An extension of eqn (10) to overlapping generations was also given by Yonezawa (1997). Equation (10) assumes that sexual and asexual contributions are independent. Predictions relaxing this assumption and extensions to more complex models were given by Yonezawa *et al*. (2000, 2004). Analytical expressions for these models of mixed sexual and asexual species were also given by Orive (1993) and Balloux *et al*. (2003) using coalescence theory.

Predictions of effective size for haplo-diploid species can generally be made by the standard formula for sex-linked genes (see review by Caballero, 1994). Some situations occur, however, where reproduction of these species is more complex than assumed by the simplest models. For example, in many eusocial Hymenoptera species, males can be produced by workers rather than only by queens. Predictions of N_e for this scenario have been developed by Nomura and Takahashi (2012).

Populations under selection

In the absence of selection or when selection acts on a non-inherited trait, the effective size is simply a function of the variance of the number of offspring per parent, as in eqn (5). However, predictions of N_e are more complicated when selection acts on an inherited trait, such as when artificial selection is carried out for a quantitative trait in animal or plant breeding, or when natural selection acts on fitness traits against deleterious mutations or in favour of advantageous ones. In these scenarios, the drift process is amplified over generations because the random associations originated in a given generation between neutral and selected genes remain in descendants for a number of generations until they are eliminated by segregation and recombination. This problem was first addressed by Robertson (1961) and later on by other authors (e.g., Wray and Thompson, 1990; Woolliams *et al*., 1993; Santiago and Caballero, 1995) for directional selection in quantitative traits. Extensions of the model were made later for populations under natural selection, linkage, overlapping generations and animal breeding schemes, as will be reviewed below.

Selection assuming unlinked genes: When selection acts on an inherited trait, changes in gene frequency at a focal neutral locus are positively correlated over generations because the selective values randomly associated with the neutral locus are not completely removed by segregation and recombination from one generation to the next. For unlinked genes and weak selection, the random association generated by sampling in a single generation is halved in consecutive generations by segregation and recombination. Therefore, the accumulative selective association has a limiting value $Q = \sum_{i=0}^{\infty}(1/2)^i = 2$ times the value of the original random association (Robertson, 1961), and the corresponding variance of the long-term contributions of copies of the neutral gene will increase by a factor Q^2. With regards to drift, the effective variance of contributions of individuals (with average 2) increases due to selection by the same factor up to $4Q^2C^2$, where the term C^2 is the genetic variance of the individual trait measures (for the quantitative trait subject to artificial selection, or fitness-related traits in the case of natural selection) relative to the mean of the trait in the population. This variance has to be added to the expected variance of random contributions not caused by selection (S_k^2) to predict the total variance of contributions. In reality, the associations are also reduced each generation to a proportion equal to the fraction of genetic variance remaining after selection (G) which, in turn, can be increased by the correlation between the selective advantages of male and female parents (r), and the series becomes

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

(Santiago and Caballero, 1995). In the case of partial selfing (or partial full-sib mating), the term r in eqn (12) should be replaced by β (the proportion of inbred matings), because the correlation between the expected selective values of males and females (r) is approximately one for inbred matings (which take place with proportion β) and approximately zero for non-inbred matings, i.e. $Q = 2/[2 - G(1 + \beta)]$.

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

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

Nomura (1999b, 2005a) showed that eqn (13), obtained by a genetic drift approach, could also be derived from an inbreeding approach by considering the variance of long-term contributions as used by Wray and Thompson (1990) and Wray *et al*. (1990), when appropriate corrections are made in the latter (see also Woolliams and Bijma, 2000).

The application to different numbers of males and females was given by Santiago and Caballero (1995). That equation, however, lacked the same covariances as the equation without selection, as shown by Nomura (1997a) and Wang (1998). For random mating (α = 0) and Poisson distribution of family sizes (S_k^2 = 2), eqn (13) reduces to the simplest expression (Robertson, 1961),

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

Equation (14) can be expressed in terms of heritability (h^2) of fertility, as shown by Nei and Murata (1966) and Nomura (2002b). Let V_k be the observed variance of family sizes, which would 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 term, $S_k^2 = V_k(1 - h^2)$, is the non-heritable component of this variance, and the second term, $4C^2 = V_k h^2$, is the heritable component. Thus, substituting these into eqn (14) yields

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

(Nei and Murata, 1966). The extension of eqn (15) to dioecious populations was developed by Nei and Murata (1966) assuming random union of gametes. A more general equation was developed by Nomura (2002b), who also suggested a form of the equation that avoids estimating the heritability,

$$N_e = \frac{4N}{2+V_k+3(cov_{k,m}+cov_{k,f})} \ , \qquad (16)$$

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

The prediction of effective population size under selection with overlapping generations was considered by Nomura (1996) and Bijma *et al*. (2000). As for the non-selection case (Hill, 1979), N_e is the same as that for populations with discrete generations having the same non-selective and selective components of variance in lifetime progeny numbers, and the same number of individuals entering the population each generation. Another interesting result is that the average age of parents in populations under selection is smaller than that in populations under no selection, as, in the 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 $Q = 2/(1 + kB_b)$ 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 effect of selection on nucleotide diversity, by Santiago and Caballero (1998) analysing the cumulative effect of selection over generations with a genetic drift approach (the Robertson effect), and by Nicolaisen and Desai (2012, 2013) using the coalescence theory. All these papers derived the same equation for the asymptotic N_e, which is a function of the haploid deleterious mutation rate U, the effect s of mutations and the length L of the whole-genome or genome segment given in Morgans,

$$N_e = Ne^{-U/(s+L/2)} \;. \tag{17}$$

This equation is the exponential version of eqn (14), $N_e = Ne^{-Q^2C^2}$, which was derived from the multiplicative fitness model assumed under background selection. Here, $C^2 = Us$ is the variance for fitness and the cumulative term, Q, for a rate of recombination c between the neutral and selected loci, is $Q_c = \sum_{i=0}^{\infty}[(1-s)(1-c)]^i \approx 1/(s+c)$ (Santiago and Caballero, 1998). If the focal neutral locus is located in the middle of the genome segment and the selected loci are uniformly scattered, the average value of the $Q_c{}^2$ terms over the segment is $Q^2 = 1/[s(s + L/2)]$. Substituting this and $C^2 = Us$ in eqn (14), we obtain eqn (17).

It is important to note that this equation predicts the magnitude of drift or inbreeding in the long-term. For the focal neutral allele, this magnitude is effectively reached after a number of generations counted since it first appeared by mutation. Until that moment, drift at the neutral locus is expected to increase with time. The increasing drift acting on neutral mutations in consecutive generations can be predicted by the partial $N_{e(t)}$ values for generation t forward in time that can be calculated using the partial cumulative terms $Q_{c(t)} = \sum_{i=0}^{t}[(1-s)(1-c)]^i = (1-e^{-(s+c)t})/(s+c)$ (Santiago and Caballero, 1998). An equivalent conclusion was reached by Nicolaisen and Desai (2012, 2013) from the point of view of the coalescent process. The consecutive $N_{e(-t)}$ values that predict the increasing probability of coalescence under selection t generations backwards in time (thus, the negative sign) reach an asymptotic value given by eqn (17) and the predictions of the partial N_e values given by both methods, forward and backward, are exactly the same for any generation t,

$$N_{e(t)} = N_{e(-t)} = Ne^{-\frac{Us}{L/2}\int_0^{L/2} Q_{c(t)}^2 dc} \;. \tag{18}$$

Illustrations of the decline in $N_{e(t)}$ over generations are given in Fig. 3 of Santiago and Caballero (1998) and Fig. 3 of Nicolaisen and Desai (2013). This shows that the distortion of coalescent genealogies under selection and the cumulative effect of selection over generations are both specular images of the same process. Moreover, the pattern of neutral variation in populations under selection can be predicted by accumulating the expected distributions of neutral mutations originated in all the previous generations with the corresponding consecutive effective sizes given by the $N_{e(t)}$ values (Santiago and Caballero, 1998; Nicolaisen and Desai, 2012, 2013). This means that the intensity of genetic drift varies over mutations that occurred at different generations backward in time. Therefore, strictly speaking, there is not a single N_e value representing the intensity of the drift process and, consequently, the amount and spectrum of variation under selection, even in populations at mutation-selection-drift equilibrium.

If mutations are advantageous (selective sweep model), predictions are generally cumbersome, because the genetic variance at selected loci is a function of the gene frequencies. Moreover, the stochastic distribution of selective sweeps over time is far away from the constant flux of variation usually assumed to simplify derivations. Wiehe and Stephan (1993) and Gillespie (2000) derived equations for the expected heterozygosity at neutral loci using a model in which recurrent favourable mutations pass quickly through the population to fixation, wiping out linked variation. The first key simplification in these derivations is that the time of fixation of favourable mutations is so short relative to the time scale of genetic drift that it can be considered as occurring instantaneously. The second simplification is that the rate of recovery of neutral variation after a selective sweep is very slow when compared with the rate of occurrence of the sweeps. The recurrent substitutions lead to a roughly constant level of neutral variation in a stochastic process that is often referred to as “genetic draft”. A simple solution for N_e can be obtained by directly applying the concept of variance of long-term contributions to an evolutionary time-scale:

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

(Gillespie, 2000), where θ is the rate of selective sweeps and y is the final frequency of the neutral copy that was initially associated with the favourable mutation when it first appeared. The frequency of this neutral copy is expected to increase to Ny copies after the sweep, and the frequency of each of the other $2N - 1$ copies is expected to be reduced to $(1 - y)$ copies. Therefore, the variance of the expected long-term contributions for a single selective sweep is about $2Ny^2$. As selective sweeps occur at a rate θ, the second term of the denominator in equation (19) is the total variance of the expected long-term contributions, that is Q^2C^2 in equation (14).

Effective population size in conservation practices

The concept of effective size is key to conservation genetics practices, as it summarises the past history of the population regarding inbreeding and genetic drift, and provides the prospects for the sustainability of the population if the current effective size is maintained in the future. The effective population size is directly related to the statistics widely used to monitor conservation breeding schemes, such as the number of genome equivalents $N_{ge} \approx N_e/t$ (Lacy, 1995), where t is the number of generations of management.

Minimising the loss of genetic variation is one of the main objectives of captive breeding programmes. This is achieved through minimising genetic drift and, therefore, maximising N_e. A classical strategy to follow is the equalisation of family sizes. By choosing one couple from each pair of parents, the variance of parental contributions is null ($S_k^2 = 0$) and, from eqn (1), $N_e \approx 2N$ (Wright, 1938; Crow, 1954), which is twice the effective size of an unmanaged population with a Poisson distribution of family size. This is known as minimal inbreeding and it is the recommended procedure for applications in germplasm collection and regeneration in plants (see, e.g., Vencovsky *et al*., 2012). However, effective sizes larger than $2N$ can actually be obtained by population subdivision (Wright, 1943; Wang and Caballero, 1999) and other strategies, as shown below. The extension of the strategy in the case of different numbers of males and females was developed by Gowe *et al*. (1959). In their selection scheme, each male contributes one son and $r_{fm} = N_f/N_m$ daughters, and each female contributes one daughter and has a probability of N_m/N_f of contributing one son. Thus, $S_{fm}^2 = (N_m/N_f)[1 - (N_m/N_f)]$ and all of the other variances and covariances in eqn (2) are zero. Substituting these into eqns (2-3) gives

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

Compared with no selection (random Poisson distribution of the number of offspring per parent, eqn 4), this scheme can increase N_e by a proportion of $(r_{fm} + 3)/(3r_{fm} + 1)$. When the female/male ratio, $r_{fm} = N_f/N_m$, is 2 , for example, N_e is increased by 71.4%.

Wang (1997c) proposed an alternative design which produces further increases in N_e of about 17% when $r_{fm} = 2$. In this scheme, among the r_{fm} females mated with each male, one is selected at random to contribute one son, and each of the remaining $r_{fm} - 1$ females contributes one 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 induced between the numbers of male and female offspring from female parents, $S_{fm,ff} = -N_m/N_f$. Substituting these terms into eqns (2-3),

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

The benefit of this scheme over that from Gowe *et al*. (1959) is decreased as r_{fm} gets larger. For sex-linked loci a benefit is also produced if males are the heterogametic sex. The above equations refer to random mating of parents. Wang (1997c) also proposed a system of non-random mating in which each male is mated with one of the groups of half-sib females that are not sisters of the male. This is a sort of population subdivision where the half sibs are like “subpopulations” and there is random migration of males and no migration of females among the “subpopulations” (see Wang and Caballero, 1999). The mating scheme can further increase N_e over that predicted by eqn (21).

Wang´s (1997c) method applies to a single generation. Sánchez *et al*. (2003) extended the method across generations to account for long-term contributions, further improving its efficiency. With the availability of pedigree or molecular marker information, a more general method, based on finding the contributions from parents to progeny which minimise the average coancestry among the progeny (minimum coancestry contributions), is the most widely proposed criterion to maintain genetic diversity (Toro and Pérez-Enciso, 1990; Ballou and Lacy, 1995; Fernández *et al*., 2003; Meuwissen, 2007). This method has been shown to minimise the variance of the long-term contributions from ancestors to descendants and, therefore, to maximise effective population size (Caballero and Toro 2000, 2002).

The above methods are all designed to reduce the variation in family sizes, the term S_k^2 in eqn (1) and its corresponding components when the numbers of males and females are different (eqn 2). It is also possible to increase the effective population size by decreasing the Mendelian segregation variance, which is represented by the *constant* term “2” in eqn (1). This latter can be achieved by the use of Marker Assisted Selection (MAS) to minimise the variation in contribution between the paternally and maternally derived genes at a locus (Wang and Hill, 2000). Thus, for example, for equal numbers of males and females and equalisation of individual contributions, N_e can be expressed as

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

where $P_{m,mf}$ ($P_{f,mf}$) is the probability that the two genes coming from the male (female) parent and contributing to their male and female progeny are copies of the same gene. By MAS it is possible to reduce these probabilities below the value of 0.5 expected under no control of Mendelian segregation, depending on the amount of marker information, the genome size and the number of marker-genotyped offspring per family, achieving values of N_e larger than $2N$. MAS can also be used in a more general framework of different numbers of males and females to minimise global genetic drift and inbreeding (Wang, 2001a).

An alternative and complementary method is to use reproductive technologies for meiosis manipulation, such as *in vitro* culture of premeiotic germ cells and microinjection of primary spermatocytes into oocytes. By using more than one gamete from a single meiosis, variation from Mendelian segregation can be partially or completely removed (Santiago and Caballero, 2001). Thus, for example, if equalisation of family sizes is carried out and the gametes from both male and female parents are managed to come from the same meiosis in each case, the resulting effective size becomes $3N$, rather than the typical $2N$.

The control of the increase in inbreeding and genetic drift in conservation programmes is 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 variance effective size, as expected. In this scenario, with various census sizes and a mixture of groups, predictions depend, however, on the generations considered (see Ryman *et al*. 1999).

For endangered species in the wild, estimating the effective population size and monitoring its changes over time is important in understanding the genetic health, evaluating the risk of inbreeding and inbreeding depression and thus the risk of extinction, assessing the effectiveness of the genetic managements (e.g. human-aided migration/relocation, habitat protection or modification), and projecting the future demographic trajectories of the populations. Simulations (Tallmon *et al*. 2010) showed that monitoring the effective size is most often a more robust means of identifying stable and declining populations than monitoring census size. If a population is detected to have a small or declining N_e, the managers of the populations should be alerted to investigate the causes and to take effective measures for reversing the course. Using noninvasive sampling (e.g. from faeces, feathers, hair, …), genetic marker data can be obtained from a population even without observing (disturbing) the animals. The data can then be analysed normally, except for accounting for the possibility of genotyping errors and allowing for a high frequency of missing data (e.g. Wang 2004), in estimating N_e. For example, in a long-term monitoring project, Kamath *et al*. (2016) sampled and genotyped (at 20 microsatellite loci) 729 Yellowstone grizzly bears (*Ursus arctos*) born in the period 1962-2010 from an isolated and well-studied population in the Greater Yellowstone Ecosystem. They used the data to study the population demographic trajectories, estimating the changes in N_e and generation interval, over this time period.

Methods for estimating the effective population size from genetic data

Given the concepts of effective size, different approaches can be used to *predict* the effective size of a population from its demographic parameters such as census size and variance of reproductive success. In parallel, different methodologies can also be developed to *estimate* the realized effective size of a population from its genetic properties revealed by genetic markers, such as temporal changes in allele frequency and linkage disequilibrium.

Quite a few methods (Schwartz *et al*., 1999; Beaumont, 2003a; Wang, 2005; Palstra and Ruzzante, 2008; Luikart *et al*., 2010; Gilbert and Whitlock, 2015) have been developed and applied to estimating N_e in widely different spatial and time scales, from ancient, past to current (parental) population sizes. Herein we focus on the effective size of the current generation or just a few generations in the past, as this time scale is the most relevant for conservation genetics (Luikart *et* *al.*, 2010) and plant and animal breeding, and is most likely to yield accurate estimates in current practices.

Heterozygosity excess

Compared with an infinitely large population at Hardy-Weinberg equilibrium, a population generated from a number of N_m male parents and a number of N_f female parents is expected to show a deficit of homozygotes and an excess of heterozygotes at a neutral locus when N_m, N_f, or both, are small. This is because male and female parents are expected to have different allele frequencies due to drift. The smaller the value of N_m or N_f, the greater the difference between paternal and maternal allele frequencies and thus the greater the excess in heterozygosity of the offspring population. There is a simple functional relationship between the N_e of the parental population and the amount of heterozygosity excess in the offspring population (e.g. Robertson, 1965; Wang, 1996a). For a Wright-Fisher ideal population except for separate sexes with N_m male and N_f female parents, the heterozygosity excess is expected to be

$$D \approx -\frac{1}{8N_m} - \frac{1}{8N_f} = -\frac{1}{2N_e}, \qquad (23)$$

where $N_e = 4N_m N_f / (N_m + N_f)$ is the effective size of the parental population given by eqn (4). The value of D is negative, indicating an excess of heterozygosity and a corresponding deficit of homozygosity. For a non-ideal population with arbitrary distributions of family sizes, (23) is still valid when N_m and N_f are replaced by N_{em} and N_{ef} respectively, the effective numbers of male and female breeders.

Eqn (23) suggests that measuring the heterozygosity excess, D, at a number of marker loci in a population yields an estimate of the parental population effective size. Pudovkin *et al.* (1996) proposed such a N_e estimator by accounting for the sampling effect,

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

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 the expected heterozygosity from the observed gene frequency $\hat{p}$, and $\widehat{H}_o$ is the observed heterozygosity. $\widehat{D}$ is calculated for each allele in a multiallelic locus and for each locus, and the average value is used in (24) (Luikart and Cornuet, 1999). The accuracy of the estimator was evaluated by Pudovkin *et al.* (1996) using simulations, and was applied to a few empirical datasets (Luikart and Cornuet, 1999). The method is simple, and is implemented in several computer programs (e.g. Zhdanova and Pudovkin, 2008; Jones and Wang, 2010; Do *et al.* 2014). However, the method has a low precision and accuracy, frequently providing infinitely large estimates of N_e for small populations. The estimator is also highly sensitive to non-random mating (e.g. population subdivision, close relative mating), which also causes deviation from Hardy-Weinberg equilibrium. Its poor performance renders it useless in applications to empirical dataset analysis, except when the actual population size is very small and marker information is ample.

Linkage disequilibrium

In a large unselected random mating population, alleles are independent within and between loci, producing Hardy-Weinberg equilibrium and linkage equilibrium. In a finite population, however, random genetic drift leads to associations between alleles at a locus and between alleles of different loci. The former results in heterozygosity excess, and the latter leads to gametic linkage disequilibrium (LD). In addition to drift, LD can also be induced by factors such as migration and direct or indirect (e.g. hitchhiking) selection (Hedrick *et al*., 1978). For neutral loci unlinked with selected loci in an isolated population under random mating, LD would come exclusively from genetic drift and can be used to estimate N_e (Hill, 1981).

A LD estimator of N_e for a random mating population at equilibrium is based on the formulation (Hill, 1981),

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

where c is the recombination rate (c = 1/2 for unlinked loci), r is the correlation of allele frequencies between two loci due to LD, and n is the sample size (number of sampled individuals). In an equilibrium population, allele frequencies at two neutral loci are expected to be uncorrelated (i.e. $E[r] = 0$), such that the expectation of squared r, $E[r^2]$, is equal to the variance of r, $V[r]$. Eqn (25) shows that $V[r]$ is composed of two distinctive parts. The first comes from genetic drift, determined by N_e and linkage c. The second comes from sampling, determined by sample size n. Using the genotypes at a number of loci of n sampled individuals, we can estimate $V[r]$, which can then be inserted in (25) to obtain an estimate of N_e if the recombination fraction c between loci is known. Note that a slightly different expression for the population $V[r]$ (i.e. the first part on the right-hand side of eqn 25) was derived by Sved (1971) from an identity by descent approach. For a dioecious population with monogamy, the right side of (25) should be increased by c / ($2N_e c(2 – c)$) (Weir and Hill, 1980).

Hill (1981) also derived the formula for the sampling variance of the estimator such that uncertainties of the N_e estimates can also be evaluated. For the case of no linkage, Waples (2006) showed by simulations that the LD estimator can seriously underestimate N_e when sample size is small. He derived empirical equations to correct for the bias caused by small sample sizes, and showed by simulations that the accuracy of the modified estimator is comparable to the temporal method described in the next section. To facilitate the applications of the LD estimator, Waples and Do (2008) published a computer program, LDNE, and further evaluated its performance in comparison with the temporal method, using simulated data (Waples and Do, 2010). They concluded that, under similar conditions in terms of marker information and the actual population size, LD estimator can yield N_e estimates that have equivalent or better qualities than the temporal estimators, except when the sampling interval of the temporal method is long.

The LD estimator is simple to calculate, and requires just a single sample of multilocus genotypes instead of two or more samples, as is with the temporal method (see below). It is especially suitable for species with a long generation interval where obtaining two samples separated by a couple of generations means many years, and for genetic monitoring (Schwartz *et al*., 2007) to track population trajectories on a yearly basis. As a result, the LD estimator has gained popularity in recent years (Palstra and Ruzzante, 2008; Luikart *et al*., 2010). However, some assumptions inherent to the LD estimators are often violated in real populations, and as a result may lead to biased N_e estimates. For example, it is assumed that LD is produced solely from the finite population size, and other confounding factors, such as non-random mating and population structure, are absent. Any departure from random mating (e.g. an excess or deficit of close relatives mating including selfing) will affect LD, and thus LD based estimates of N_e. Waples *et al*. (2014) evaluated the effect of age structure on LD estimators, and found that LD calculated from mixed-age adult samples is overestimated and thus N_e is underestimated in all of 21 simulated species with different life tables. Similarly, the LD in a subpopulation is reduced by a constant and high rate of immigration and elevated by a low rate of immigration, compared with that of an isolated population of the same N_e. Therefore, as observed by Waples and England (2011) in their simulation study, LD calculated from a sample from a subpopulation leads to an overestimate or an underestimate of local N_e when immigration rate is high or low, respectively. In the former case, the estimated local N_e converges to the global N_e of the entire population (Waples and England, 2011).

LD is highly dependent on the recombination rate between loci (Hill, 1981). Pairs of closer linked loci have higher LD, and thus provide better information about N_e (suitably defined in time scale) if the linkage information among loci is known. Although Hill (1981) derived his LD estimator of N_e allowing for an arbitrary level of linkage, and he advocated the use of linked markers, most often unlinked markers are used in practice because either truly unlinked markers are used or potentially linked markers are used but their linkage relationship is unknown. LD of markers of different recombination rates sheds light on the effective size of the population in different time periods in the past (Wang, 2005). Quite a few methods (Hayes *et al*. 2003; Barbato *et al*. 2015; Mezzavilla and Ghirotto 2015; Saura *et al*. 2015) have been developed to exploit the LD information from many densely spaced markers on a chromosome segment in inferring the N_e at different time points in the past.

Temporal changes in allele frequency

For an infinitely large population under Hardy-Weinberg equilibrium, both allele frequencies and genotype frequencies remain unchanged over time. In reality, these frequencies never stay constant, and change systematically due to the forces of mutation, selection and migration, stochastically due to the random force of genetic drift, or both. In the absence of the action of all of the systematic forces in a population, any observed allele frequency change must come solely from genetic drift and can thus be used to infer the rate of drift or the N_e of the population. Based on this logic, Krimbas and Tsakas (1971) proposed to measure allele frequency changes at a number of marker loci between two temporally separated samples of individuals, and thereby to estimate the N_e of the population during the sampling interval. This so-called 'temporal method' was subsequently developed by many others in both (allele frequency) moment (e.g. Nei and Tajima, 1981; Pollak, 1983; Waples, 1989) and likelihood (e.g. Williamson and Slatkin, 1999; Anderson *et al*., 2000; Wang, 2001b; Berthier *et al*., 2002; Beaumont, 2003b; Laval *et al*., 2003) approaches.

Moment estimators calculate a standardized variance in the temporal changes of allele frequency, *F*, from marker genotypes in two temporally spaced samples. *F* is essentially similar to Wright's F_{ST}, the differences being that *F* measures the temporal differentiation for the same population and it also includes sampling effect. There are a few *F* estimators (e.g. Nei and Tajima, 1981; Pollak, 1983) available, the one being widely applied was derived by Nei and Tajima (1981). This estimator is calculated by

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

for a locus with *k* alleles, where x_i and y_i are the observed frequencies of allele *i* in the first and second samples respectively. For multiple loci, $\hat{F}$ is obtained by averaging single locus estimates. The expectation of $\hat{F}$ depends on the sampling schemes (sampling with or without replacements), and is a function of N_e and sample sizes to account for genetic drift and sampling effects. Solving the expectation equation of $\hat{F}$ for N_e yields the temporal estimate of the (harmonic) mean N_e during the sampling period (Nei and Tajima, 1981; Waples, 1989).

Moment estimators rely on the summary statistic, *F*, which is simple to calculate. However, they do not use the full allele frequency information and are thus less accurate than the probabilistic methods. The latter, likelihood or Bayesian, are much more complicated in statistical modelling and in computation. In general, temporal methods provide good estimates of N_e when it is not large, sampling interval is not too short (e.g. one generation), and the assumptions of the methods are satisfied, using a typical set of 10-20 microsatellites. Likelihood methods generally have higher accuracy and precision than moment methods, especially for markers with rare alleles, as verified by several extensive simulations (e.g. Wang, 2001b; Berthier *et al.*, 2002; Tallmon *et al.*, 2004). They are however much more computationally demanding than moment methods, which complete an analysis almost instantly. Recently, the computational efficiency of likelihood methods has been improved substantially by Hui and Burt (2015), using a hidden Markov algorithm and applying continuous approximations to allele frequencies and transition probabilities. The new method can deal with N_e values as high as several millions, and is implemented in an R package called NB.

A constraint on the applications of temporal approaches is the requirement of at least two samples taken at one or preferably more generations apart. The longer this sampling interval, t, is, the stronger the drift signal will be in the temporal data, and the more accurate the N_e estimate will become. The extent of drift is proportional to t and is inversely proportional to N_e. For the same population and the same sampling intensity (in terms of the number of markers, number of samples, and number of individuals per sample), increasing the sampling interval t could improve the accuracy of the temporal methods tremendously, as has been repeatedly verified by simulations (e.g. Nei and Tajima, 1981). Moment estimators have especially low accuracy when t is small (say, $t < 3$) because of the weak drift signal and also because of the approximations made to the estimators. In practice, it is unfortunately difficult or unrealistic to increase t, especially for long-lived species having a long generation interval.

Compared with other N_e estimating approaches, the temporal approach makes fewer assumptions and is more robust to some complications (realities) in real populations. For example, the approach is robust to population structure. It applies to a single unsubdivided population under non-random mating (including selfing), and to a subdivided population when sampling is representative and the aim is to estimate the N_e of the entire subdivided population rather than that of a single subpopulation. It is also robust to age structure in a population with overlapping generations, so long as the sampling interval t is large (e.g. Nei and Tajima, 1981). When t is small, however, Waples and Yokota (2007) showed by simulations that typical sampling regimes (sampling only newborns, only adults, and all age classes in proportions) result in biased N_e estimates. Jorde and Ryman (1995) developed a moment estimator of N_e applicable to populations with overlapping generations. They derived an age-structure correction factor, which, when applied to the standard moment estimator for populations with discrete generations, leads to unbiased N_e estimates for populations with overlapping generations. Unfortunately, however, the correction factor is a function of numerous age-specific survival rates and age-specific reproduction rates of the focal population. These rates are usually unknown. In fact, once all these rates are known for a population, the N_e of the population can be calculated from standard N_e prediction equations (e.g. Felsenstein, 1971; Hill, 1972, 1979) without the need of genetic data. A method to calculate the N_e of a population with overlapping generations from its demographic parameters has been implemented in an R-package Neff (Grimm *et al*. 2016).

The standard temporal approach for a single unsubdivided population was also extended to estimate the N_e of a subpopulation which is connected to other subpopulations by gene flow (Wang and Whitlock, 2003). Although both drift and immigration change allele frequencies of a subpopulation, the detailed patterns of the changes are different between drift and immigration. Using temporal samples from a focal subpopulation and a sample from a large source population (the island-mainland model) or from two focal subpopulations (the island-island model), a moment estimator and a likelihood estimator can yield joint estimates of N_e and migration rates, m. Simulations showed (Wang and Whitlock, 2003) that both moment and likelihood estimators gave reasonably good estimates of N_e and m under typical sampling intensities. However, no estimators are available for the more general case of multiple ($n > 2$) subpopulations. Part of the difficulty is with the number, n^2, of parameters to be jointly estimated, including n effective sizes and $n(n - 1)$ migration rates. More work is badly needed in this direction as spatial and temporal genotype data are becoming easy to collect thanks to the rapid developments in molecular technologies.

One of the assumptions in the temporal approach is the absence of selection so that any change in allele frequency comes solely from drift and thus indicates the effective size of the population. For most marker loci, the assumption is valid, especially for a small population over a short sampling interval of just a few generations. However, over a long period, some loci could be affected by adaptive selection or purging selection and their allele frequencies could change faster or slower than those of neutral loci unaffected by selection. Allele frequencies at neutral loci could also evolve faster or slower because of linkage disequilibrium with those under selection. The temporal methods have been extended to estimate N_e of a population and the selection coefficient, s, of a locus from time-series data of allele frequencies (e.g. Bollback *et al*., 2008; Mathieson and McVean, 2013; Foll *et al*., 2015). These methods are usually Bayesian, based on hidden Markov models to explain the observed allele frequency changes due to drift and selection. How well these methods perform has yet to be checked, perhaps by a simulation study.

Relatedness and relationship

The pattern of genetic relatedness or relationship between individuals in a population has a direct functional relationship with the inbreeding effective size of the population (Wang, 2009). Two individuals taken at random from a population with a smaller N_e will have a higher probability of sharing the same father, mother, or both. More generally, the mean and variance in pairwise relatedness within a generation are expected to increase with decreasing N_e. Based on this logic, Nomura (2008) proposed a method to use the increase in average coancestry between two consecutive generations to estimate N_e. He showed by simulations that his coancestry method is more biased, but more precise than the heterozygosity excess method. The overall accuracy (measured by mean squared errors) of the two methods is similar. A major problem which causes the bias of the method, as recognized by Nomura (2008), is that some non-sib pairs must be selected from a sample of individuals to act as reference in estimating the mean coancestry. The selection of non-sib pairs is difficult and somewhat subjective, because it is now well-known that classifying dyads into even well separated relationship categories, e.g. full sibs, half sibs, parent offspring, and unrelated, from pairwise relatedness estimates is highly error prone (e.g. Blouin *et al.*, 1996). Although many marker-based pairwise relatedness estimators are unbiased, they have high sampling errors with no exceptions (Wang, 2014).

A more robust and powerful method is to estimate the frequencies of half-sib (sharing a single parent) and full-sib (sharing both parents) dyads, Q_{HS} and Q_{FS}, in a sample taken at random from a single cohort of a population (Wang, 2009; Waples and Waples, 2011). Wang (2009) derived a formula of N_e in terms of half- and full-sib frequencies, using both an inbreeding and a drift approach,

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

The equation has the parameter α as in eqn (5), so that the N_e for a population under non-random mating (e.g. partial selfing) can be estimated. While Q_{HS} and Q_{FS} can be estimated from a sibship assignment analysis of the multilocus genotypes (Wang and Santure, 2009; Jones and Wang, 1010), α can be estimated from the same data with a F_{ST}-like approach (Wang, 2009). Alternatively, α can be assumed to be zero for an outbred population when marker genotype frequencies do not deviate significantly from those expected under Hardy-Weinberg equilibrium. The difficulty comes from the estimation of the numbers of breeding males, N_m, and females, N_f, because sibship analysis generally makes no distinctions between paternal and maternal sibships from autosomal marker data, except in some specific situations (Wang, 2009). However, as detailed in Wang (2009), the 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 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 value.

There are several advantages of this sibship approach compared with other single-sample 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 multiple pieces of information may potentially allow for a better delineation of the process and thus yield a more accurate estimate of N_e. Tallmon *et al*. (2008) proposed to use approximate Bayesian computation (ABC) to estimate N_e from a sample of microsatellite genotypes. Their method, implemented in a computer program ONESAMP, calculates and uses eight summary statistics which are known to have functional relationships with N_e from population genetics theory or simulations. These statistics include, among others, the number of alleles per locus, expected heterozygosity, linkage disequilibrium, Wright's F_{IS}, the mean and variance of multilocus homozygosity. In essence, the ABC approach simulates populations of different N_e and tries to find the N_e value that yields the same or similar summary statistics to those calculated from the real data. Tallmon *et al*. (2008) demonstrated this ABC approach by analysing an introduced increasing population of ibex *Capra ibex*.

It is arguable that the ABC approach uses more information than other approaches. On one hand, it uses multiple sources of information such as heterozygosity, number of alleles, and LD. However, on the other hand, for each source of information, it uses a summary statistic rather than the full information that is used by the probability methods (likelihood or Bayesian). Furthermore, it is unclear how these different summary statistics should be optimally weighted, given that these statistics are, apparently, highly correlated and may reflect the inbreeding and drift processes of different time scales. For a population changing in size, these different summary statistics are relevant for N_e in different time scales. For example, F_{IS} (like heterozygosity excess) is pertinent to the parental N_e, LD implicates N_e in the past few or more generations (depending on the linkage of the markers), while the number of alleles can be determined by the ancient N_e many generations (in the order of N_e or $1/u$, whichever is smaller, where u is mutation rate) ago. So far an extensive simulation study to compare the accuracy of ABC and other approaches is lacking, but is urgently needed.

Discussion

Since the seminal work of Wright (1931, 1933), great progress has been made on the pivotal population genetics parameter, N_e, in its concepts (e.g. inbreeding, variance, eigenvalue effective sizes, etc.), its predictions for various species under different mating systems and population structures, and its estimation methodologies using various marker information. In parallel, estimates of N_e, from both demographic and genetic data, have been made for many populations in the past 30 years, thanks to the rapid developments in both molecular technologies and statistical and computational methodologies.

Much work has been done to predict the effect of selection on N_e. However, developing useful predictive models on the effect of selection acting on an inherited trait remains difficult. The reason is that the impact of linked genes propagates over a number of generations, resulting in a long-term effect that is difficult to combine in a simple equation with parameters referred to only one generation time. Coalescence theory runs into similar difficulties in predicting N_e, because the probabilities of coalescence for consecutive generations are not independent under selection on an inherited trait. In addition to the variation of N_e across generations, there is also variation over the genome. It is nowadays quite clear that there is a significant heterogeneity in levels and patterns of genetic variation across the genome caused by selection (Charlesworth, 2009; Gossmann *et al*., 2011), which complicates the inferences of N_e.

Another important remaining problem about selection is the interrelationship between N_e and genetic variation. Most equations of N_e are linear functions of the census size N where genetic variation of the selected trait is an independent variable. However, genetic variation itself depends on N_e. Ignoring this fact is irrelevant for some purposes, but is troublesome in some models of closely linked loci. This reciprocal dependence is in the basis of the Hill-Robertson effect (Hill and Robertson, 1966) and Muller´s ratchet (Haigh, 1978), both being different aspects of the same issue, an additional reduction of genetic variance due to genetic drift induced by selection.

Demographic estimation of N_e can be made by application of the predictive equations reviewed here when information on census sizes, variances of progeny numbers, type of mating system, and other demographic data, are available. The lack of these data and the increasing availability of genetic markers make the estimation of N_e through genetic data to be, however, the leading procedure. Most factors affecting the populations in real situations imply a reduction of the effective size relative to the census size. In fact, the observed ratio N_e / N has been found to be about 10%-20% (Frankham, 1995; Palstra and Fraser, 2012) on average in meta-analyses across many species and populations. Overall, these figures are in agreement with theoretical expectations obtained from some of the predictive equations presented in this review when fluctuations in population size are considered (Vucetich *et al*., 1997). However, this average N_e / N ratio may be an overestimate, as marine species are under-represented in these meta-analyses and can have extremely low N_e / N ratios.

Each N_e estimation method with genetic data is based on a certain population genetics model and has a number of assumptions. It is important to realize that, when these assumptions are violated, which is unfortunately the rule rather than the exception in the real world, an estimation method may yield invalid or biased estimates of N_e. For example, most methods reviewed herein assume an isolated random mating population with discrete generations. Yet in practice such populations are rare. The robustness of different methods has not been fully investigated.

Another important issue is the interpretation of the estimates obtained from a certain method. First, what is the spatial scale relevant to a N_e estimate? Is it the effective size of the local population from which samples are taken, or that of the metapopulation of which the sampled local population is a part? Apparently, the local and metapopulation effective sizes are very different in quantities and in applications. Local and metapopulation N_es signify the intensities of inbreeding and genetic drift processes at the local and global levels, respectively. A small local N_e but a large global N_e (i.e. many small interconnected populations) and a large local N_e but a small global N_e (i.e. a few large interconnected populations) have different ecological, evolutionary, and conservation genetics implications.

Second, what is the temporal scale relevant to a N_e estimate (Waples 2005)? Is it the N_e of the sampled population, of the parental population, of an ancestral population or the average N_e over several/many previous generations? The heterozygosity excess method and sibship methods estimate the parental population N_e, temporal methods estimate the (harmonic) average N_e over the generations in the sampling interval, while LD and ABC methods estimate the average N_e over an unspecified number of previous generations (Wang, 2005). Of course the temporal scale becomes irrelevant for a population with a constant unchanging demography. In practice, however, a natural population never stays the same.

Third, does the estimate refer to inbreeding or variance effective size? The question is irrelevant for the case of an isolated population with a constant demography or an incompletely subdivided (i.e. with migration) population of constant size, as the inbreeding (N_{eI}) and variance (N_{eV}) effective sizes are the same. Otherwise, however, N_{eI} and N_{eV} can be dramatically different. A decreasing (increasing) population will always have a N_{eI} greater (smaller) than N_{eV}, because the former and latter depend on the stochastic processes in the parental and the offspring generations, respectively (Crow and Kimura, 1970; Caballero, 1994; Wang and Caballero, 1999). Although usually unspecified in the original work describing the estimators, they estimate N_{eI}, N_{eV} or a combination of both. The heterozygosity excess and sibship method estimate N_{eI}, while the temporal methods estimate N_{eV}. It is unclear what the LD and ABC methods estimate, but they likely estimate a combination of both N_{eI} and N_{eV}.

Fourth, are the estimators unbiased and accurate when their underlying assumptions are met and are violated? Most estimators are not fully evaluated for their performances and statistical behaviours by simulation studies, especially those that are computationally intensive. In measuring the overall accuracy of an estimator, both precision and bias should be considered, and better incorporated into a single measurement, such as mean squared errors. It is better to measure the mean squared errors of $1/(2N_e)$ rather than N_e, because the latter can be infinitely large and, more importantly, it is invariably $1/(2N_e)$ rather than N_e that enters a population genetics equation (Wang and Whitlock, 2003). The dominating factor in determining accuracy is precision and bias when marker information is scarce and ample, respectively.

More work is also needed in developing estimators that make fewer restrictive assumptions and thus are more widely applicable to real populations. A common challenge to the current estimators is population genetic structure, including age structure (i.e. overlapping generations) and spatial structure (i.e. population subdivision with migration). Blindly applying an estimator developed for a single isolated population with discrete generations under random mating to an age- or space-structured population may yield unpredictable results.

Most current estimators have good performance for a population with a small N_e, partly because the drift or inbreeding signal is strong. For a large population with N_e in the thousands or more, drift and inbreeding in the population is weak and is thus difficult to pick up by the typical sampling intensity (say, ~100 individuals, 10-20 microsatellites). With the rapid developments of genotyping (sequencing) technology, an increasing number of studies use many genome-wide markers to investigate population structure and demography. Use of thousands of SNPs may compensate for the weak signal of stochasticity in a large population, and thus may allow for a good estimate of large N_e. Future work is needed in evaluating current methods to estimate large N_e using many markers. It can be reasonably assumed that, for an accurate estimate of N_e (say, a narrow 95% confidence interval of roughly $[0.5N_e, 2N_e]$) by any efficient method, a sample size of individuals, of the total number of effective alleles across loci, or of both in the order of N_e might be required.

A related issue with large N_e and many markers is the computational efficiency of likelihood or Bayesian estimators. Efficient algorithms, such as that proposed recently for the likelihood temporal estimator (Hui and Burt, 2015), are urgently needed to deal with big datasets. Another option is to exploit the modern multicore and multi-cpu computers and to parallel computer program codes using techniques such as MPI (Message Passing Interface) and OpenMP.

In some applications, the parameter, N_e, is all that is required in describing the current and predicting the future genetic variation in a population, and the demographic details that determine N_e are irrelevant. In other applications such as the conservation management of endangered species, however, both N_e and the demographic details of the population are useful in designing the most effective management to maintain the genetic diversity (Wang, 2009). When a population is estimated to have a small N_e and thus is prone to the loss of genetic variation, we may ask what the causes are. Is the small N_e caused by a small number of breeders, by a large variance in reproductive success among breeders, by an unbalanced sex ratio, etc? Different causes imply different optimal management strategies. The utility of future N_e estimators could be improved if they provided joint estimates of N_e and important demographic quantities such as variance of reproductive success.

Acknowledgements

We are grateful to Xav Harrison, the editor and five anonymous referees for helpful comments on previous versions of this MS. AC was funded by Ministerio de Economía y Competitividad (CGL2012-39861-C02-01), Xunta de Galicia (GPC2013-011) and Fondos Feder: “Unha maneira de facer Europa”.

References

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.

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

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

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

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

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

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

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

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

Caballero A, Toro MA (2000). Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genet Res* **75**: 331–343.

Caballero A, Toro MA (2002). Analysis of genetic diversity for the management of conserved subdivided populations. *Conserv Genet* **3**: 289–299.

Caballero A, Wei M, Hill WG (1996b). Survival rates of mutant genes under artificial selection using individual and family information. *J Genet* **75**: 63–80.

Charlesworth B (2001). The effect of life-history and mode of inheritance on neutral genetic variability. *Genet Res* **77**: 153–166.

Charlesworth B (2009). Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet* **10**: 195–205.

Charlesworth B (2013). Background selection 20 years on: The Wilhelmine E. Key 2012 invitational lecture. *J Hered* **104**: 161–171.

Charlesworth B, Charlesworth D (2010). *Elements of Evolutionary Genetics*. Roberts & Co, Greenwood Village, Colorado, USA.

Crow JF (1954). Breeding structure of populations. II. Effective population number. In *Statistics and Mathematics in Biology*, pp. 543-556. Iowa State COIL Press. Ames, Iowa.

Crow JF, Morton NF (1955). Measurement of gene frequency drift in small populations. *Evolution* **9**: 202–214.

Crow JF, Denniston C (1988). Inbreeding and variance effective population numbers. *Evolution* **42**: 482–495.

Crow JF, Kimura M (1970). *An Introduction to Population Genetics Theory*. Harper & Row, New York.

Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR (2014). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Mol Ecol Res* **14**: 209-214.

Engen S, Lande R, Saether BE (2005). Effective size of a fluctuating age-structured population. *Genetics* **170**: 941–54.

Evans BJ, Charlesworth B (2013). The effect of nonindependent mate pairing on the effective population size. *Genetics* 193: 545–556.

Ewens WI (1979). *Mathematical Population Genetics*. Springer, Berlin.

Falconer D, Mackay TFC (1996). *Introduction to Quantitative Genetics*. Longman, Harlow.

Felsenstein J (1971). Inbreeding and variance effective numbers in populations with overlapping generations. *Genetics* **68**: 581–597.

Fernández J, Toro MA, Caballero A (2003). Fixed contributions designs versus minimization of global coancestry to control inbreeding in small populations. *Genetics* **165**: 885–894.

Fernández J, Toro MA, Caballero A (2004). Managing individuals' contributions to maximize the allelic diversity maintained in small, conserved populations. *Conserv Biol* **18**: 1–10.

Fisher, RA (1965). *The Theory of Inbreeding*. Academic Press, New York, Second edition.

Foll M, Shim H, Jensen JD (2015). WFABC: a Wright–Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. *Mol Ecol Res* **15**: 87–98.

Frankham RD (1995). Effective population size / adult population size ratios in wildlife: a review. *Genet Res* **66**: 95-107.

Frankham R, Ballou JD, Briscoe DA (2010). *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge, UK.

Ghai GL (1969). Structure of populations under mixed random and sib mating. *Theor Appl Genet* **39**: 179–182.

Gilbert KJ, Whitlock MC. (2015). Evaluating methods for estimating local effective population size with and without migration. *Evolution* **69:** 2154-2166.

Gillespie JH (2000). Genetic drift in an infinite population: the pseudo hitchhiking model. *Genetics* **155**: 909–919.

Gossmann TI, Woolfit M, Eyre-Walker A (2011). Quantifying the variation in the effective population size within a genome. *Genetics* **189**: 1389–1402.

Gowe RS, Robertson A, Latter BDH (1959). Environment and poultry breeding problems. 5. The design of poultry control strains. *Poul Sci* **38**: 462–471.

Grimm A, Gruber B, Hoehn M, Enders K, Henle K (2016). A model-derived short-term estimation method of effective size for small populations with overlapping generations. *Methods in Ecology and Evolution* (in press).

Haigh J (1978). The accumulation of deleterious genes in a population–Muller's Ratchet. *Theor Popul Biol* **14**: 251–267.

Haldane JBS (1924). A mathematical theory of natural and artificial selection. H. The influence of partial self-fertilisation, inbreeding, assortative mating, and selective fertilisation on the composition of Mendelian populations, and on natural selection. *Proc. Camb. Phil. Soc., (Biol. Sci.), (later Biol. Rev.)* **1**: 158–163.

Haldane JBS (1939). The equilibrium between mutation and random extinction. *Ann Eugen* **9**: 400–405.

Hare MP, Nunney L, Schwartz MK, *et al*. (2011). Understanding and estimating effective population size for practical application in marine species management. *Conserv Biol* **25:** 438-449.

Hayes BJ, Visscher PM, McPartlan HC, Goddard ME (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res* **13**: 635–643.

Hedrick P, Jain S, Holden L (1978). Multilocus systems in evolution. *Evol Biol* **11**: 101-184.

Hill WG (1972). Effective size of populations with overlapping generations. *Theor Popul Biol* **3**: 278–289.

Hill WG (1979). A note on effective population size with overlapping generations. *Genetics* **92**: 317–322.

Hill WG (1981). Estimation of effective population size from data on linkage disequilibrium. *Genet Res* **38**: 209–216.

Hill WG, Robertson A (1966). The effects of linkage on limits to artificial selection. *Genet Res* **8**: 269–294.

Hudson RR, Kaplan NL (1995). Deleterious background selection with recombination. *Genetics* **141**: 1605–1617.

Hui TYJ, Burt A (2015). Estimating effective population size from temporally spaced samples with a novel, efficient maximum-likelihood algorithm. *Genetics* **200**: 285-293.

Jones OR, Wang J (2010). COLONY: a program for parentage and sibship inference from multilocus genotype data. *Mol Ecol Res* **10**: 551-555.

Jorde PE, Ryman N (1995). Temporal allele frequency change and estimation of effective size in populations with overlapping generations. *Genetics* **139**: 1077–1090.

Kamath PL, Haroldson MA, Luikart G, Paetkau D, Whitman C, Manen FT (2015). Multiple estimates of effective population size for monitoring a long-lived vertebrate: an application to Yellowstone grizzly bears. *Mol. Ecol.* **24**, 5507-5521.

Kimura M, Crow JF (1963a). The measurement of effective population number. *Evolution* **17**: 279–288.

Kimura M, Crow JF (1963b). On the maximum avoidance of inbreeding. *Genet Res* **4**: 399–415.

Krimbas CB, Tsakas S (1971). The genetics of *Dacus oleae* V. Changes of esterase polymorphism in a natural population following insecticide control: selection or drift? *Evolution* **25**: 454–460.

Lacy R (1995). Clarification of genetic terms and their use in the management of captive populations. *Zoo Biol* **14**: 565–578.

Laporte V, Charlesworth B (2002). Effective population size and population subdivision in demographically structured populations. *Genetics* **162**: 501–519.

Laporte V, Cuguen J, Couvet D (2000). Effective population sizes for cytoplasmic and nuclear genes in a gynodioecious species: the role of the sex determination system. *Genetics* **154**: 447–458.

Laval G, SanCristobal M, Chevalet C (2003). Maximum-likehhood and Markov chain Monte Carlo approaches to estimate inbreeding and effective size from allele frequency changes. *Genetics* **164**: 1189–1204.

Li CC (1976). *First Course in Population Genetics*. Boxwood, Pacific Grove, CA.

Luikart G, Cornuet JM (1999). Estimating the effective number of breeders from heterozygote excess in progeny. *Genetics* **151**: 1211–1216.

Luikart G, Ryman N, Tallmon DA, Schwartz MK, Allendorf FW (2010). Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conserv Genet* **11**: 355-373.

Mathieson I, McVean G (2013). Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. *Genetics* **193**: 973–984.

Mezzavilla M, Ghirotto S (2015). Neon: An R package to estimate human effective population size and divergence time from patterns of linkage disequilibrium between SNPs. *J Comput Sci Syst Biol* **8**: O37-O44.

Meuwissen THE (2007). Operation of conservation schemes. In: Oldenbroek, K. (Ed.), Utilisation and Conservation of Farm Animal Genetic Resources. Wageningen Academic Publishers, Wageningen, The Netherlands, pp. 167–193.

Nagylaki T (1992). *Theoretical Population Genetics*. Springer, Berlin.

Nei M, Murata M (1966). Effective population size when fertility is inherited. *Genet Res* **8**: 257–260.

Nei M, Tajima F (1981). Genetic drift and estimation of effective population size. *Genetics* **98**: 625–640.

Nicolaisen LE, Desai MM (2012). Distortions in genealogies due to purifying selection. *Mol Biol Evol* **29**: 3589–3600.

Nicolaisen LE, Desai MM (2013). Distortions in genealogies due to purifying selection and recombination. *Genetics* **195**: 221–230.

Nomura T (1996). Effective size of selected populations with overlapping generations. *J Anim Breed Genet* **113**: 1–16.

Nomura T (1997a). On the effective size of nonrandom mating populations under selection. *Jpn J Biometrics* **18**: 1–11.

Nomura T (1997b). Effective population size for a sex-linked locus in populations under selection. *Math Biosci* **142**: 79–89.

Nomura T (1997c). Prediction of effective population size in open nucleus breeding systems. *J Anim Breed Genet* **114**: 333–347.

Nomura T (1998a). Effective population size in supportive breeding. *Conserv Biol* **13**: 670–672.

Nomura T (1998b). Inbreeding and effective size of populations under selection. Proceedings of the MAFF International Workshop on Genetic Resources Animal Genetic Resources II, pp. 5–24, National Institute of Agrobiological Resources, Tsukuba.

Nomura T (1999a). Inbreeding coefficient and effective size of finite populations with a specified proportion of sib-mating. *Jpn J Biometrics* **83**: 485–489.

Nomura T (1999b). On the methods for predicting the effective size of populations under selection. *Heredity* **83**: 485–489.

Nomura T (2000). Effective population size under marker assisted selection. *Jap J Biometrics* **21**: 1–12.

Nomura T (2002a). Effective size of populations with unequal sex ratio and variation in mating success. *J Anim Breed Genet* **118**: 297–310.

Nomura T (2002b). Effective size of populations with heritable variation in fitness. *Heredity* **89**: 413–416.

Nomura T (2005a). Developments in prediction theories of the effective size of populations under selection. *Anim Sci J* **76**: 87–96.

Nomura T (2005b). Effective population size under random mating with a finite number of matings. *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.

Robertson A (1965). The interpretation of genotypic ratios in domestic animal populations. *Anim Prod* **7**: 319–324.

Rönnegård L, Woolliams JA (2003). Predicted rates of inbreeding with additive maternal effects. *Genet Res* **82**: 67–77.

Ryman N, Laikre L (1991). Effects of supportive breeding on the genetically effective population size. *Conserv Biol* **5**: 325–329.

Ryman N, Jorde PE, Laikre L (1995). Supportive breeding and variance effective population size. *Conserv Biol* **9**: 1619–1628.

Ryman N, Jorde PE, Laikre L (1999). Supportive breeding and inbreeding effective number: Reply to Nomura. *Conserv Biol* **13**: 673–676.

Sánchez L, Bijma P, Woolliams JA (2003). Minimizing inbreeding by managing genetic contributions across generations. *Genetics* **164**: 1589–1595.

Santiago E, Caballero A (1995). Effective size of populations under selection. *Genetics* **139**: 1013–1030.

Santiago E, Caballero A (1998). Effective size and polymorphism of linked neutral loci in populations under directional selection. *Genetics* **149**: 2105–2117.

Santiago E, Caballero A (2001). Application of reproductive technologies to the conservation of genetic resources. *Conserv Biol* **14**: 1831–1836.

Saura M, Tenesa A, Woolliams JA, Fernández A, Villanueva B. (2015). Evaluation of the linkage-disequilibrium method for the estimation of effective population size when generations overlap: an empirical case. *BMC Genomics* **16:** 922.

Schwartz MK, Tallmon DA, Luikart G (1999). Using genetics to estimate the size of wild populations: many methods, much potential, uncertain utility. *Anim Conserv* **2**: 321–323.

Schwartz MK, Luikart G, Waples RS (2007). Genetic monitoring as a promising tool for conservation and management. *TREE* **22**: 25-33.

Sved JA (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor Pop Biol* **2**: 125-141.

Tallmon DA, Luikart G, Beaumont MA (2004). Comparative evaluation of a new effective population size estimator based on approximate Bayesian computation. *Genetics* **167**: 977–988.

Tallmon DA, Koyuk A, Luikart G, Beaumont MA (2008). COMPUTER PROGRAMS: onesamp: a program to estimate effective population size using approximate Bayesian computation. *Mol 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.

Wang J (2004). Sibship reconstruction from genetic data with typing errors. *Genetics* **166**: 1963-1979.

Wang J (2005). Estimation of effective population sizes from data on genetic markers. *Phil Trans R Soc B* **360**: 1395–1409.

Wang J (2009). A new method for estimating effective population sizes from a single sample of multilocus genotypes. *Mol Ecol* **18**: 2148–2164.

Wang J (2014). Marker-based estimates of relatedness and inbreeding coefficients: an assessment of current methods. *J Evol Biol* **27**: 518-530.

Wang J, Caballero A (1999). Developments in predicting the effective size of subdivided populations. *Heredity* **82**: 212–226.

Wang J, Hill WG (2000). Marker assisted selection to increase effective population size by reducing Mendelian segregation variance. *Genetics* **154**: 475–489.

Wang J, Ryman N (2001). Genetic effects of multiple generations of supportive breeding. *Conserv Biol* **15**: 1619-1631.

Wang J, Santure AW (2009). Parentage and sibship inference from multilocus genotype data under polygamy. *Genetics* **181***:* 1579-1594.

Wang J, Whitlock MC (2003). Estimating effective population size and migration rates from genetic samples over space and time. *Genetics* **163**: 429-446.

Wang J, Brekke P, Huchard E, Knapp LA, Cowlishaw G (2010). Estimation of parameters of inbreeding and genetic drift in populations with overlapping generations. *Evolution* **64**: 1704–1718.

Wang Y, Pollak E (2002). The effective number of population that varies cyclically in size. II. Overlapping generations. *Math Biosci* **179**: 161–181.

Waples RS (1989). A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* **121**: 379–391.

Waples RS (2005) Genetic estimates of contemporary effective population size: to what time periods do the estimates apply? *Mol Ecol* 14: 3335-3352.

Waples RS (2006). A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conserv Genet* **7**: 167–184.

Waples RS, Do C (2008). LdNe: a program for estimating effective population size from data on linkage disequilibrium. *Mol Ecol Res* 8: 753–756.

Waples RS, Do C (2010). Linkage disequilibrium estimates of contemporary Ne using SNPs and highly polymorphic molecular markers: an evaluation of precision and bias. *Evol Appl* **3**: 244-262.

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.

Wright S (1931). Evolution in Mendelian populations. *Genetics* **16**: 97–159.

Wright S (1933). Inbreeding and homozygosis. *Proc Natl Acad Sci USA* **19**: 411–420.

Wright S (1938). Size of population and breeding structure in relation to evolution. *Science* **87**: 430–431.

Wright S (1939). *Statistical genetics in relation to evolution. Exposés de Biométrie et de Statistique Biologique*. Herman & Cie, Paris, France.

Wright S (1943). Isolation by distance. Genetics 28: 114-138.

Wright S (1969). *Evolution and the Genetics of Populations, vol. 2, The Theory of Gene Frequencies*. Chicago: University of Chicago Press.

Yonezawa K (1997). Effective population size of plant species propagating with mixed sexual and asexual reproduction system. *Genet Res* **70**: 251–258.

Yonezawa K, Kinoshita E, Watano Y, Zentoh H (2000). Formulation and estimation of the effective size of stage-structured populations in *Fritillaria camtschatcensis*, a perennial herb with a complex life history. *Genetics* **54**: 2007–2013.

Yonezawa K, Ishii T, Nagamine T (2004). The effective size of mixed sexually and asexually reproducing populations. *Genetics* **166**: 1529–1539.

Zhdanova OL, Pudovkin AI (2008). Nb_HetEx: a program to estimate the effective number of breeders. *J Hered* **99**: 694-695.