

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Publications, Agencies and Staff of the U.S.  
Department of Commerce

U.S. Department of Commerce

---

2010

## Spatial-temporal stratifications in natural populations and how they affect understanding and estimation of effective population size

Robin Waples

NOAA, robin.waples@noaa.gov

Follow this and additional works at: <https://digitalcommons.unl.edu/usdeptcommercepub>

---

Waples, Robin, "Spatial-temporal stratifications in natural populations and how they affect understanding and estimation of effective population size" (2010). *Publications, Agencies and Staff of the U.S. Department of Commerce*. 472.

*Department of Commerce*. 472.

<https://digitalcommons.unl.edu/usdeptcommercepub/472>

This Article is brought to you for free and open access by the U.S. Department of Commerce at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Publications, Agencies and Staff of the U.S. Department of Commerce by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

THE STATE OF THE FIELD

# Spatial-temporal stratifications in natural populations and how they affect understanding and estimation of effective population size

ROBIN S. WAPLES

NOAA Fisheries, Northwest Fisheries Science Center, 2725 Montlake Blvd. East, Seattle, WA 98112, USA

## Abstract

The concept of effective population size ( $N_e$ ) is based on an elegantly simple idea which, however, rapidly becomes very complex when applied to most real-world situations. In natural populations, spatial and temporal stratifications create different classes of individuals with different vital rates, and this in turn affects (generally reduces)  $N_e$  in complex ways. I consider how these natural stratifications influence our understanding of effective size and how to estimate it, and what the consequences are for conservation and management of natural populations. Important points that emerge include the following:

1. The relative influences of local vs metapopulation  $N_e$  depend on a variety of factors, including the time frame of interest.

2. Levels of diversity in local populations are strongly influenced by even low levels of migration, so these measures are not reliable indicators of local  $N_e$ .

3. For long-term effective size, obtaining a reliable estimate of mutation rate is the most important consideration; unless this is accomplished, estimates can be biased by orders of magnitude.

4. At least some estimators of contemporary  $N_e$  appear to be robust to relatively high (approximately 10%) equilibrium levels of migration, so under many realistic scenarios they might yield reliable estimates of local  $N_e$ .

5. Age structure probably has little effect on long-term estimators of  $N_e$  but can strongly influence contemporary estimates.

6. More research is needed in several key areas: (i) to disentangle effects of selection and drift in metapopulations connected by intermediate levels of migration; (ii) to elucidate the relationship between  $N_b$  (effective number of breeders per year) and  $N_e$  per generation in age-structured populations; (iii) to perform rigorous sensitivity analyses of new likelihood and coalescent-based methods for estimating demographic and evolutionary histories.

*Keywords:* age structure, contemporary  $N_e$ , long-term  $N_e$ , metapopulation, migration

*Received 6 January 2010; revision received 15 April 2010; accepted 16 April 2010*

## Introduction

Effective population size ( $N_e$ ) is one of the most elegantly simple concepts in all of biology. Whereas the census size ( $N$ ) is of primary importance for shaping demographic and ecological processes (birth and death rates; competition; predation), a population's response to evolutionary

forces is determined by its effective size. The elegant part of the  $N_e$  concept is the idea that a single number can directly determine the rate of one of the four evolutionary forces (random genetic drift) and capture essential information about the relative importance of the other three (mutation, migration, and selection).

However, virtually everything else about effective population size is complex. The point of reference for evaluating effective size is a hypothetical 'ideal' population of constant size that is closed to immigration and has

Correspondence: Robin Waples, Fax: (206) 860-3335; E-mail: robin.waples@noaa.gov

discrete generations, equal sex ratio, random mating, and random variation in reproductive success. These conditions are never completely satisfied in natural populations. Wright (1931, 1938) showed how the concept of  $N_e$  can account for skewed sex ratio, non-random variation in reproductive success, and cyclical fluctuations in population size. Subsequent evaluations have demonstrated that the effective size concept can also be applied to various non-ideal situations (age-structured populations; complex mating systems; migration; captive-wild systems; etc.). However, most of these applications also required simplifying assumptions that are not realistic for natural populations. In particular, in most natural populations, spatial and temporal stratifications create different classes of individuals with different mean vital rates (age-specific survival and fecundity), and this in turn affects  $N_e$  in complex ways.

In this article, I consider how spatial and temporal factors, and their interaction, influence our understanding of the concept of effective size. I will be particularly concerned with practical applications to real-world problems in evolution, conservation, and management. The first part of the article considers how these spatial and temporal factors affect  $N_e$  and the  $N_e/N$  ratio; next, I consider how these same factors influence estimates of these key parameters.

### Concepts and definitions

In the first half of the 20th Century, emergence of the field of population genetics produced major advances in our understanding of evolutionary processes. Population size ( $N$ ) plays an important role in virtually all such analyses. To facilitate these analyses, the concept of a Wright-Fisher (WF) 'ideal' population was developed. In a WF population, each of the  $N$  individuals has an equal opportunity to contribute genes to the next generation. Conceptually, this can be accomplished if each individual contributes equally to an infinite pool of gametes, which unite at random to form  $N$  individuals for the next generation. The WF process thus envisions binomial sampling, which is well-characterized mathematically, and assuming WF dynamics allows one to model genetic drift as a binomial process that involves sampling  $2N$  genes (for a diploid species) repeatedly across generations.

The WF population also plays another important role in providing an ideal yardstick against which to compare the evolutionary behaviour of actual populations, which rarely if ever exactly correspond to WF dynamics. Wright introduced the concept of a population's effective size as the size of an ideal WF population that would have the same rate of random genetic drift as the population in question. Subsequently, Crow (1954) distinguished two flavours of effective size that measure different aspects of

genetic drift: variance  $N_e$  (which reflects the rate of allele frequency change) and inbreeding  $N_e$  (which reflects the rate of increase in identity by descent). The two flavours of  $N_e$  are the same in closed populations of constant size and will not be distinguished below unless necessary for clarity.

Conceptually, then,  $N_e$  is defined in terms of rates of genetic drift, and the point of reference is a continuum of ideal populations of different numbers of ideal individuals; for every real population, there is an hypothetical ideal population of  $N_e$  individuals that will have the same rate of genetic drift as the population in question. The effective size concept can be made operational with reference to key demographic data:  $\bar{k}$  (mean) and  $V_k$  (variance) in the number of gametes contributed by each parent to the next generation. For example, for a random mating monoecious population with random selfing (the original WF model), inbreeding effective size is given by (Crow & Denniston 1988; Equation 1; Caballero 1994, Equation 22):

$$N_e = \frac{\bar{k}N - 1}{\bar{k} - 1 + V_k/\bar{k}}. \quad (1)$$

An ideal population is constant in size, which requires  $\bar{k} = 2$ , leading to

$$N_e = \frac{2N - 1}{1 + V_k/2} = \frac{4N - 2}{2 + V_k}. \quad (2)$$

Random union of the  $2N$  gametes leads to binomial variance in reproductive success, so in a population, the expected value of  $V_k$  is  $2(N-1)/N$ . By definition, these processes produce  $N_e = N$  in a WF population, as can be verified by inserting the binomial variance into Eqn (2). However, in real populations, various factors typically conspire to ensure that individuals do not have an equal opportunity to contribute genes to the next generation, with the result that (in general)  $N_e < N$ .

Here, I introduce a general way of thinking about this problem that considers biological stratification, model constraints, and process variation (wobble). In the basic WF model, a single panmictic population is envisioned and all individuals have similar properties and expectations for survival and reproduction (i.e. all individuals are interchangeable). In the real world, spatial and/or temporal stratification occurs naturally through biological and physical properties, and this stratification creates different expectations of demographic and genetic parameters for different groups of individuals. Stratification also creates additional layers of complexity, and as a result, theoretical (and sometimes computer) models often impose constraints on evolutionary behaviour by making simplifying assumptions about the stratification process. In particular, for simplicity and to make analyses tractable, many models ignore random variations associ-

ated with the stratification process. The term ‘wobble’ is used here to refer to random fluctuations in population parameters that occur naturally but are constrained or ignored in many models.

### Spatial-temporal stratifications

#### *Separate sexes*

An obvious natural form of stratification occurs if the species under consideration has separate sexes. In that case, the WF condition that each individual has an equal opportunity to contribute to the next generation can only be met if the sex ratio is unity; otherwise, members of the less numerous sex will have higher average reproductive success because half the genes for the next generation must pass through each sex. Unless the sex ratio is 1:1, therefore, the sexes will have different mean  $\bar{k}$ , overall  $V_k$  will be greater than the binomial, and  $N_e$  will be less than  $N$ . In a population that is otherwise ideal, a simple formula captures the consequences of unequal sex ratio on effective size (Wright 1938):

$$N_e = \frac{4MF}{M+F}, \quad (3)$$

where  $M$  and  $F$  are the numbers of males and females, respectively, and  $M + F = N$ .

Based on Eqn (3), it is easy to show that with separate sexes,  $N_e = N$  only if  $M = F$ . Therefore, extending the WF concept of an ideal population to a species with separate sexes requires the additional constraint that the numbers of males and females are exactly equal. Note that this is only possible if  $N = M+F$  is an even number, so sexually outbreeding populations with an odd number of individuals cannot be ideal. This requirement also constrains  $V_k$ , because the WF model requires that  $V_k$  within each sex must be binomial. As the binomial variance increases with  $N$ , and as the numbers of each sex are only half the total size  $N$ , overall variance in reproductive success is constrained by the assumption that  $M = F$ . This can be shown quantitatively as follows. If  $\bar{k} = 2$  and  $M = F = N/2$ , then assuming binomial sampling implies that for each sex,

$V_k = 2 \frac{(N/2 - 1)}{N/2} = \frac{2(N - 2)}{N}$ , and inserting this into Eqn (2) yields

$$N_e = \frac{2N - 1}{1 + (N - 2)/N} = N \frac{2N - 1}{2N - 2} \approx N + \frac{1}{2},$$

which agrees with results reported by others using different approaches (Caballero 1994; Balloux 2004). That is, adding the provision for separate sexes and the constraint that  $M = F$  causes the effective size of an ideal

population with separate sexes to be approximately one-half an individual larger than the comparable value for a monoecious population.

It is easy to force sex ratio to be equal in an analytical or computer model, and in some real-world cases, this can also be achieved by manipulation (for example, in a captive breeding program where specific matings can be arranged). More generally, however, in natural populations, there is no reason to expect the sex ratio to be exactly 1:1 at any given point in time, even if that is the central tendency. Even if every newborn has an equal probability of being male or female, random variation will ensure that the realized sex ratio seldom is exactly equal. This problem was considered by Waples & Do (1994; Appendix), who showed that in a population that is otherwise ideal but has the sex of each newborn chosen randomly,  $E(N_e) = N - 1$ . That is, random processes involving separate sexes in natural populations will in general slightly reduce  $N_e$  compared to what it would be for monoecious populations, whereas the opposite conclusion (separate sexes slightly increase  $N_e$ ) is reached if one artificially constrains the original WF model to ensure exactly equal sex ratio. Although both effects are small unless  $N$  is small, this example illustrates the point that ignoring or artificially constraining natural wobble can produce misleading conclusions.

#### *Age structure*

A key assumption of the WF model is that generations are discrete, so reproduction in a single season also represents an individual’s lifetime reproductive output. Felsenstein (1971) and Hill (1972) considered  $N_e$  in species with overlapping generations using models that are somewhat complementary. Hill’s is more general, as it makes no particular assumptions about variance in reproductive success and provides a formula that is analogous to similar formulas for species with discrete generations (Hill 1972):

$$N_e \approx \frac{4N_1 T}{V_k + 2}. \quad (4)$$

In this formula (note the strong similarity to Eqn (2)),  $N_1$  is the number of newborns entering the population each year,  $T$  is generation length, and  $V_k$  is lifetime variance in reproductive success among the individuals in a cohort. Felsenstein’s model is more limited in some respects as it assumes Poisson variance of reproductive success within each age class; however, his model can be linked more directly to demographic data (age-specific probabilities of survival and reproduction) and, unlike Hill’s method, can accommodate populations that are growing or declining exponentially. Hill’s model produces the same results

as Felsenstein's under the Poisson-variance assumption (Johnson 1977; Charlesworth 1980).

Using the current paradigm, we can see that age structure represents a biological stratification of the population into winners and losers in the sweepstakes of survival and reproduction. Consider a species that has  $N_1$  newborns each year and lives to a maximum age  $L$  and has age-specific survival and reproductive rates of  $s_x$  and  $b_x$ , respectively, where  $x = 1, 2, \dots, L$  is age. Whereas the standard WF model has one class of individuals (all of whom have the same expectations for key parameters) and the two-sex variation has two classes of individuals (males and females), iteroparous species with overlapping generations have  $L$  classes of individuals based on age at death ( $j$ )—those that die after reproducing at age 1, after reproducing at age 2, etc. Mean lifetime reproductive success thus increases with age at death:  $\bar{k}_j = \sum_{x=1}^j b_x$ . Thus, although each individual within a class has the same expected lifetime reproductive success, individuals that die at different ages can vary widely in their mean contributions to the next generation. As a result, variance in lifetime reproductive success across an entire cohort is larger than under the WF model, even if the variance within each class is no greater than random.

The standard WF model assumes constant population size, and Hill and Felsenstein extended this assumption to include stable age composition in species with overlapping generations (see Gaggiotti and Vetter 1999 for an application to a marine fish). The assumption of stable age composition implies exactly the same number of individuals in each age class each year. This, of course, is unlikely to be true in nature, where demographic and environmental stochasticity can be expected to produce random fluctuations around expected values. However, relatively little study has been given to the consequences for  $N_e$  of ignoring this source of wobble. Nunney (2002), Vitalis *et al.* (2004) and Waples (2002a) used theoretical and numerical methods to evaluate  $N_e$  in fluctuating populations of semelparous species with variable age at maturity (like Pacific salmon (*Oncorhynchus* spp.) and annual plants with seed banks), and Waples (2006) showed that results depend heavily on the method of population regulation. Engen *et al.* (2005) considered the problem in a more general way for age-structured species in variable environments and found that  $N_e$  is inversely related to generation length and demographic/genetic variance but is independent of environmental variance.

#### Changes in $N_e$ over time

Genetic drift in its various forms (allele frequency change; increase in inbreeding in finite populations; loss of heterozygosity) occurs at a rate that is inversely

proportional to effective size, and this non-linear relationship has important consequences for considering 'average' behaviour across multiple generations. For example, in the absence of mutation, the expected fractional loss of heterozygosity in a single generation is  $1/(2N_e)$ , and the cumulative loss over  $t$  generations is given by  $H_t = H_0 \prod_{i=1, t} [1 - 1/(2N_{e(i)})]$ , where  $H_0$  and  $H_t$  are heterozygosities at generations 0 and  $t$ , and  $N_{e(i)}$  is effective size in the  $i^{\text{th}}$  generation. A good approximation to this is given by  $H_t = H_0 [1 - 1/(2\tilde{N}_e)]^t$ , where  $\tilde{N}_e$  is the harmonic mean of the  $N_{e(i)}$ , and this forms the basis for the rule of thumb that long-term  $N_e$  is determined by the harmonic mean of the effective sizes over the time period in question (see Wright 1938). Because the harmonic mean is affected much more strongly by small values,  $\tilde{N}_e$  is generally less than  $\bar{N}_e$ , and often a great deal less.

Even with constant  $N$ ,  $N_e$  can vary if mating structure (e.g. selfing rate) changes over time. Furthermore, as discussed in the next section, patterns of connectivity between subpopulations can also affect metapopulation  $N_e$ , which can therefore change if the degree of subdivision changes over time.

#### Metapopulations

*Effective size of a metapopulation.* The standard WF model assumes a single, closed, panmictic population, whereas most natural populations exist as part of a metapopulation that involves some sort of demographic and genetic linkages among component subpopulations. Natural biological and physical processes stratify the metapopulation as a whole into smaller, localized units, with the result that a given individual is more likely to mate with an individual from the same unit than with one from another unit. Wright's (still widely used) island model was the first formal attempt to capture important population genetic consequences of this type of stratification. In the finite version of this model, the metapopulation as a whole (census size  $N_T$  individuals) consists of  $n$  subpopulations, each with a constant number  $N$  of ideal individuals. Each generation, each subpopulation contributes the same fraction  $m$  of its individuals to a global migrant pool and receives back the same number of randomly chosen immigrants.

As by definition  $N_e = N$  in every subpopulation, the global census size  $N_T$  is also given by  $N_T = nN = nN_e$ . Wright showed that under these conditions, the global (metapopulation) effective size ( $N_{eT}$ ) is a simple function of  $N_T$  and  $F_{ST}$ , a measure of differentiation among subpopulations:

$$N_{eT} \approx \frac{N_T}{1 - F_{ST}}. \quad (5)$$

This equation shows that, in Wright's island model, global effective size will be larger than the sum of the subpopulation  $N_e$ s if there is any differentiation among subpopulations ( $F_{ST} > 0$ ). Based on the well-known approximation for the island model that  $F_{ST} \approx 1/(1 + 4mN_e)$ , this relationship can also be expressed in terms of the  $N_e/N$  ratio (Waples 2002b):

$$\frac{N_{eT}}{N_T} \approx 1 + \frac{1}{4mN_e}.$$

These results, however, depend heavily on several key assumptions in the island model that significantly constrain natural process variation. First, instead of simply assuming that a single panmictic population of  $N_T$  individuals is ideal (with binomial  $V_k = \bar{k}(N_T - 1)/N_T$ ), the model now must assume that each of the  $n$  subpopulations of  $N$  individuals is ideal, implying that in each subpopulation  $V_k = \bar{k}(N - 1)/N$ . Wright's model thus constrains  $V_k$  to be lower than it would for a single ideal population of the same total size, with the result that global  $N_e$  would be increased even without any genetic differentiation among subpopulations. [In the extreme, with each subpopulation at fixed size  $N = 2$  non-selfing diploids, every individual in the entire metapopulation is constrained to produce exactly two offspring each generation.] Second, the requirement that each subpopulation have exactly  $N$  individuals each generation is more restrictive than the stipulation that the total metapopulation size is constant. The rate of genetic drift is a function of  $1/N_e$ , and if  $N_e$  varies across space and time, the genetic consequences will be a function of the harmonic mean  $N_e$ , which is smaller than the arithmetic mean. This means that under Wright's model, multigenerational  $N_e$  in a typical subpopulation is higher than it would be if population size were allowed to vary randomly around the same mean. Furthermore, mandating constant  $N$  makes it impossible for subpopulations to go extinct, which can have profound effects on metapopulation  $N_e$ . For example, in the extreme case of  $F_{ST} = 1$  (each subpopulation fixed for one of two allelic variants), metapopulation  $N_e$  is infinitely large according to Eqn (5). This occurs in Wright's model because each subpopulation is immortal, so when fixation occurs by chance, those alleles are 'frozen' in place and can never be lost. Finally, the island model assumes that  $m$  is constant over time and identical across all subpopulations, but that is not a realistic depiction of natural systems (see Fraser *et al.* 2007 for an empirical example). Because genetic differentiation is an inverse function of  $m$ , variable  $m$  has non-linear effects on genetic diversity that differ from what would occur with constant  $m$  and the same mean. Whitlock (1992) showed that allowing temporal variation in demographic

parameters can have large effect on levels of genetic diversity in a metapopulation.

The island model thus involves a number of assumptions that considerably constrain process variation inherent to most natural systems, and the same is true for many related metapopulation models (e.g. 1- and 2-dimensional stepping stone) that assume ideal subpopulations of constant size. Authors who have considered more realistic models of population subdivision have made the following points:

1. Metapopulation  $N_e$  is typically much higher if migration is assumed to follow the 'migrant pool' model (each subpopulation draws migrants randomly from the same global migrant pool, as in the island model; see Wade 1975) than under the 'propagule pool' model (migrants come primarily or entirely from a single, perhaps nearby, subpopulation).
2. Frequent extinctions/recolonizations can dramatically reduce metapopulation  $N_e$ , particularly under 'propagule pool' dynamics.
3. Other factors that stratify the metapopulation into groups with different mean expectations for survival and reproduction (e.g. temporal or spatial variations in productivity) also tend to reduce metapopulation  $N_e$ .
4. In general, under most realistic scenarios, population subdivision probably reduces, rather than enhances, global effective size.

More detailed treatments of these issues can be found in Whitlock & Barton (1997), Hedrick & Gilpin (1997), Nunney (1999), Wang & Caballero (1999), Waples (2002b), Tufto & Hindar (2003), and Whitlock (2004). Recent work continues to illustrate the complex challenges to understanding effective size in connected systems. For example, Wares & Pringle (2008) showed  $N_e$  is reduced in advective systems (generated, for example, by many ocean currents) that involve unidirectional transport of individuals away from natal locations. Other studies found that ephemeral midge metapopulations show unexpectedly high levels of genetic diversity at regional scales (Berendonk *et al.* 2009) and that migration restores genetic variation in cyclical vole populations despite repeated bottlenecks (Berthier *et al.* 2006). In dendritic systems, both theoretical and empirical results indicate that hierarchical structure and asymmetric gene flow can promote maintenance of high levels of genetic diversity and therefore high metapopulation  $N_e$  (Morrissey & de Kerckhove 2009). Kuparinen *et al.* (2009) found that not only was metapopulation  $N_e$  in Atlantic salmon less than the sum of the subpopulation  $N_e$ s, but also that

it was strongly affected by the population producing the most emigrants.

*Selection and local adaptation in a metapopulation.* Although the consequences of genetic drift are predictable (at least in a statistical sense), natural selection and migration can take so many different forms that jointly analysing effects of these three evolutionary forces is quite challenging. Recent simulation results (Whitlock 2003; Whitlock & Gomulkiewicz 2005; Vuilleumier *et al.* 2008), which build upon older theoretical models, have established the following:

1. Local adaptation increases the probability of global fixation in a metapopulation, even for alleles favoured in only one locality;
2. Effects of local adaptation are diminished as subpopulation extinction rate increases, but alleles that become fixed do so more quickly;
3. Analytical approximations perform well for the weak migration limit (mutations are fixed or lost between migration events, in which case, behaviour is driven more by drift and hence local  $N_e$ ) and the strong migration limit (large  $N$  and  $m$ , low mutation rate, in which case, the effects of subdivision are reduced and the system behaves more like a single large population), but not for many intermediate (and more realistic) scenarios.
4. Local processes are relatively more important in stepping stone models than in the island model.

It is well known that in a closed population, selection is more efficient if  $N_e$  is large (for the same reason that casinos inevitably make money in the long run), while random drift can overwhelm selection if  $N_e$  is small. For metapopulations, a key question is, Does the relative importance of selection and drift depend more on local  $N_e$  or metapopulation  $N_e$ ? As we see in the next section, (Estimation of  $N_e$ ), this question is important because the nature and magnitude of migration, as well as experimental design, determines whether genetic estimators are more sensitive to local  $N_e$  or metapopulation  $N_e$ . Unfortunately, the studies mentioned previously were not designed to address this question directly.

Fortunately, however, some unpublished modelling results (Vuilleumier *et al.* in revision) are more directly relevant, and important points from this work can be summarized as follows:

For a two-deme system ( $n = 2$ ), with one fixed at  $N_1 = 100$  ideal individuals, if selection favours a local

allele that originates in a single copy in deme 1 and  $m = 0.05$ , then

1. Probability of global fixation ( $P$ ) declines as size of  $N_2$  increases, but not as fast as would occur if size of  $N_1$  were increased (provided the initial frequency of the favoured allele is held constant).
2. Source-sink dynamics enhance the decline in  $P$  as size of  $N_2$  increases.
3. If selection coefficients in the two demes are the same ( $s_1 = s_2$ ),  $P \sim 2s$  regardless of other parameters

For an  $n$ -deme system with one focal subpopulation of size  $N_1 = 100$  and the other  $(n-1)$  subpopulations all of size  $N_i$  such that total metapopulation size is constant, then if  $m = 0.05$  and selection favours an allele in the focal subpopulation but is neutral elsewhere:

1.  $P$  is much higher in the stepping stone than island model, but time to fixation is also longer;
2.  $P$  is higher with large  $n$  and small  $N_i$ ;
3. Under the island model (balanced migration), fragmentation has little effect on  $P$ .

Collectively, these results indicate that under a variety of conditions that should be plausible for natural populations, local (subpopulation)  $N_e$  can play a strong role in mediating the effectiveness of local adaptation in a metapopulation.

### Estimation of $N_e$

Although effective size is conceptually defined in terms of the rate of genetic change,  $N_e$  is operationally defined using demographic parameters. For a particular generation in a particular population, if  $\bar{k}$  and  $V_k$  are known or can be measured, realized  $N_e^*$  can be calculated directly using a formula like Eqn (1). The resulting value can be interpreted as follows: if data for that population could be collected over an increasingly larger sample of neutral, independent gene loci, the mean increase in identity by descent for the generation in question would converge on the value  $1/(2N_e^*)$ . In most natural populations, however, the necessary demographic data are difficult to collect, so it has become increasingly common to estimate  $N_e$  from molecular markers that are sensitive to various consequences of genetic drift. Detailed summaries and discussions of various genetic methods for estimating  $N_e$  can be found elsewhere (Leberg 2005; Wang 2005; Luikart *et al.* 2010); here, I focus more narrowly on how spatial

and temporal stratifications in natural populations affect genetic estimates of effective size.

### Long-term estimates

Although conceptually  $N_e$  is defined in terms of rates of genetic drift, if one assumes that evolutionary processes have been at least quasi-stable over long periods of time, it is possible to estimate long-term effective size based on the products of evolution—the number and distribution of different alleles segregating in a species or population. For example, evolutionary theory indicates that the equilibrium amount of neutral genetic variability (heterozygosity, nucleotide diversity, number of alleles) that can be maintained in a population is a function of  $\theta = 4N_e\mu$ , where  $\mu$  is the mutation rate. If  $\theta$  or a related measure can be estimated from a sample of individuals, and if mutation rate can be estimated from empirical data, simple rearrangement of the theoretical relationship provides a basis for estimating long-term  $N_e$ :

$$\hat{N}_e = \hat{\theta} / (4\hat{\mu}). \quad (6)$$

Effective size estimated this way is often referred to as the coalescent effective size because it explicitly incorporates mutation and because as  $N_e$  becomes large, the ‘complicated details of a population collapse to the Kingman Coalescent’ (Wakeley & Sargsyan 2009). Eqn (6) makes it apparent that reliability of estimates of long-term  $N_e$  depend not only on the validity of the theoretical relationship between  $N_e$ ,  $\theta$ , and  $\mu$ , but also on accuracy and precision of estimates of the latter two parameters. An estimate of uncertainty associated with long-term  $\hat{N}_e$  can be obtained by incorporating uncertainty in these other parameters (see Ruegg *et al.* 2010 for an example).

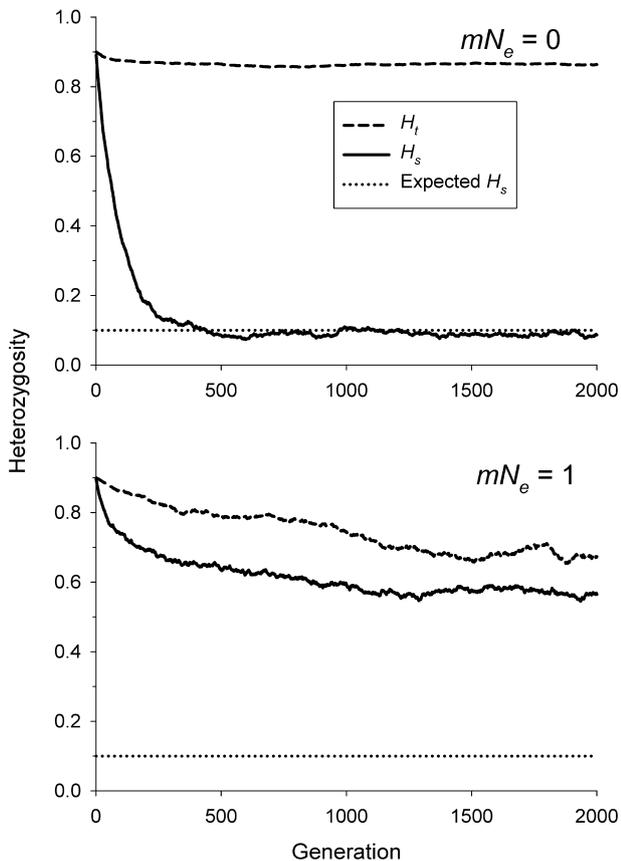
In general, given the ready availability of numerous molecular markers, sampling error in estimating  $\theta$  should not represent much of a limitation to estimates of long-term  $N_e$ . Although small samples can bias estimates of genetic diversity, the effects are rather modest and can be accounted for quantitatively (e.g. Nei 1978; Weir & Cockerham 1984). However, the sample-size adjustment will not account for biases that result from samples that are not random or representative (e.g. if they include a disproportionate number of closely related individuals). A recent article by Biro & Dingemans (2009) (‘Sampling bias from animal personality’) reminds us how difficult it is to obtain a truly random sample from a natural population.

It is apparent from the form of Eqn (6) that long-term estimates of  $N_e$  will only be as accurate as the estimate of mutation rate. For example, if the estimate of  $\mu$  is 10× higher or lower than the true mutation rate, then the

estimate of long-term  $N_e$  will also be an order of magnitude too low or too high, respectively. This is a crucial issue because estimates of mutation rates across different taxa and across different parts of the genome vary by several orders of magnitude (Ho *et al.* 2005; Henn *et al.* 2009). Furthermore, estimates for the same gene in the same species can also vary dramatically depending on the time frame over which the estimates are calculated (Howell *et al.* 2003).

An important assumption underlying Eqn (6) is that the population in question is closed to immigration. For this reason, long-term estimates of  $N_e$  have generally been applied only to species that can be considered to be roughly panmictic over large geographic areas (e.g. Avise *et al.* 1988; Ruegg *et al.* 2010). Effects of population structure on measures of genetic diversity are illustrated in Fig. 1. This figure depicts simulated data for two different migration scenarios involving an island model metapopulation:  $mN_e = 0$  (complete isolation; Panel 1A) and  $mN_e = 1$  (one migrant per generation; Panel 1B). In both panels, two measures of genetic diversity are shown:  $H_s$  = mean expected (Hardy-Weinberg) heterozygosity within each subpopulation and  $H_t$  = expected heterozygosity across the metapopulation as a whole. Panel 1A shows a wide disparity between  $H_s$  and  $H_t$ , as  $H_s$  equilibrates near the value expected for an isolated subpopulation with  $N_e = 50$ , while  $H_t$  reflects the larger  $N_e$  for the ‘metapopulation’ as a whole. Panel 1B shows a very different result: even migration as rare as one individual per generation is sufficient to ensure that each subpopulation contains a large fraction of the global diversity. Note in Fig. 1 that  $H_t$  is higher with complete isolation than for  $mN_e = 1$ , even though both scenarios involve a total of  $N_T = 20 \times 50 = 1000$  ideal individuals. This occurs for the reason described previously and exemplified in Eqn (5): the unrealistic assumptions of Wright’s island model constrain natural fluctuations and lead to higher global  $N_e$  with stronger isolation.

It is easy to visualize from Fig. 1 some of the potential pitfalls population structure poses for estimating long-term  $N_e$  and how they depend on the particular question of interest. For example, if one is interested in estimating long-term  $N_e$  for a local subpopulation that is only meaningful if the population is essentially completely isolated, as even a small amount of migration is sufficient to ensure that the level of genetic diversity in a local subpopulation is more indicative of metapopulation  $N_e$  than local  $N_e$  (Panel 1B). If one is interested in estimating local  $N_e$  for an isolated subpopulation but (inadvertently) has taken a sample that includes more than one subpopulation, the result also can be considerable upward bias (Panel 1A); however, if the subpopulations are truly isolated, they should be strongly differentiated and this scenario should be detectable using other genetic



**Fig. 1** Relationship between mean within-population expected heterozygosity ( $H_s$ ) and expected heterozygosity for a metapopulation as a whole ( $H_t$ ) as a function of level of gene flow ( $mN_e$ ) and time since initialization. Straight dotted line shows expected value of  $H_s$  for a local subpopulation; thicker curved lines show data for simulated Wright-Fisher populations (EasyPop; Balloux 2001). Simulations used 20 subpopulations of 50 ideal individuals each in an island model; each of 20 neutral gene loci had a maximum of 10 allelic states and a mutation rate of  $5 \times 10^{-4}$ , and the first generation was initiated with the maximal diversity option.

techniques, such as clustering methods. Conversely, if one is interested in estimating global/species-wide/metapopulation  $N_e$  but sampling has occurred in only a single subpopulation, the result will be a serious underestimate if the subpopulations are completely isolated. However, as the amount of migration increases,  $H_s$  and  $H_t$  converge and the estimate will not be materially different whether sampling is only local or from the metapopulation as a whole.

An alternative way to address potential biases from population structure is to use a model that explicitly accounts for migration. For example, the equilibrium model of Beerli & Felsenstein (2001) estimates  $\theta$  for each subpopulation individually, as well as the directional

(and potentially asymmetric) migration rates between pairs of subpopulations. In this model, the local estimates of  $\theta$  can be viewed as an index of local (rather than global) effective size, scaled by mutation rate. Similarly, the F model (reviewed by Gaggiotti & Foll 2010) can provide population-specific  $F_{ST}$  values by explicitly acknowledging the reality that most populations have different effective sizes and migration/immigration rates.

These methods for estimating long-term  $N_e$  explicitly assume that evolutionary processes have been stable long enough for levels of genetic diversity to have reached a dynamic equilibrium. Changes in  $N_e$  can substantially affect long-term estimates of effective size. For example, Fig. 1A shows that it took several hundred generations after complete isolation for heterozygosity to decline from a level expected for a population with  $N_e = 1000$  to the level expected for  $N_e = 50$ , and restoration of diversity by mutation following a bottleneck can take an even longer time. This is an important caveat, as the increasing pace of anthropogenic disturbance to natural ecosystems often leads to population fragmentation and decline. Conversely, some species (e.g. house sparrows, white-tailed deer, zebra mussels, purple loosestrife) thrive in human-altered environments and have recently experienced large population expansions.

In some cases, these non-equilibrium situations can be exploited to extract information about historical bottlenecks and other evolutionary changes. For example, some newer coalescent methods of genetic data analysis can explicitly account for complex changes in demographic history. For example, the model developed by Storz & Beaumont (2002) estimates the history of expansion and contraction in a closed population based on a single sample of genes, while that developed by Hey & Nielsen (2004) attempts a similar feat with interacting groups of populations.

#### *Estimates of contemporary $N_e$*

Short-term or contemporary  $N_e$  (roughly speaking, the effective size that encompasses the period over which data are collected) is important because it provides insights into local demographic and evolutionary processes. Estimates of contemporary  $N_e$  can be based on either a single sample or two samples separated in time (the temporal method). Again, detailed descriptions of the individual methods can be found elsewhere: here, I will only distinguish among them as needed. The most commonly used methods to estimate contemporary  $N_e$  assume closed populations with discrete generations, whereas most species are age structured and function as systems of interconnected populations. Therefore, I will focus this section on considering how spatial and

temporal stratifications in natural populations affect estimates of contemporary  $N_e$ .

*Local vs metapopulation  $N_e$ .* In the preceding section, we saw that over evolutionary time scales and under equilibrium conditions, even small amounts of migration are sufficient to ensure that most of the existing variation for the metapopulation as a whole is segregating within each local subpopulation. Similarly, over large numbers of generations, temporal estimates of effective size converge on the global (metapopulation) rather than local  $N_e$  (Wang and Whitlock 2003). In contrast, estimates of contemporary  $N_e$  primarily reflect local evolutionary processes that have occurred in recent generations. Although these topics have not received as much attention in the published literature as they deserve, some recent results indicate that estimators of contemporary  $N_e$  can be much more robust to migration than are estimators of long-term  $N_e$ . For example, under an equilibrium island model (where genetic distinctiveness of immigrants is inversely proportional to migration rate), the single-sample  $N_e$ -estimator based on linkage disequilibrium (LD) shows surprisingly little effect of migration unless the rate is fairly high ( $m \sim 10\%$  or higher; Phillip England, personal communication, November 2009). Similarly, over short time periods in an equilibrium island model, temporal estimates that assume closed populations are fairly robust to  $m \sim 10\%$  (Wang and Whitlock 2003; Gordon Luikart, personal communication, January 2010). These results suggest that under many natural conditions, at least some of the most widely used contemporary estimators can provide a robust estimate of local (subpopulation)  $N_e$ . An important caveat, however: anthropogenic changes to natural ecosystems can create strongly non-equilibrium demographic and evolutionary processes, and a pulse influx of substantial numbers of genetically divergent immigrants could strongly affect most or all estimators of  $N_e$ . Wang and Whitlock (2003) showed that in the temporal method, effects on  $\hat{N}_e$  can differ substantially between equilibrium and non-equilibrium scenarios, and they also developed a modified temporal method that can explicitly account for some types of migration and produce separate estimates of  $m$  and  $N_e$ . Application to real-world data sets remains challenging, however (see meta-analyses by Palstra & Ruzzante 2008), and biologically implausible estimates can occur if model assumptions are not met (Hoffman *et al.* 2004).

*Age structure.* The main problem posed by age structure is obtaining a suitable sample from an entire generation. If generations are discrete, this is conceptually straightforward, although logistical constraints in the field might pose challenges. For age-structured populations, however, a researcher interested in estimating  $N_e$  has to

address two questions related to sampling: (i) What mix of individuals of different ages/life stages would constitute a random sample from the generation as a whole? and (ii) Is it feasible to collect the necessary samples in the field? Some insight into Question 1 is provided by Felsenstein's (1971) suggestion that the way to calculate parametric allele frequency at a given point in time in an age-structured population is to genetically sample all individuals and weight them by their reproductive value. This scheme places greater emphasis on genes carried by reproductively active individuals, who are responsible for producing most of the next generation. However, this approach requires detailed demographic information, and the weighting scheme increases the variance of the estimator, which presents difficulties in practical applications (Waples & Yokota 2007).

Researchers faced with these challenges have generally adopted one of three tactics. In some special cases, it has been possible to adjust the standard temporal model to accommodate age structure. Jorde & Ryman (1995) showed that it is possible to estimate generational  $N_e$  in stable, age-structured populations if samples from consecutive cohorts are available (or can be constructed by sorting individuals by age). However, this method also requires considerable demographic information and has not seen widespread use. Waples (1990) developed a modified temporal method that accounts for species with a Pacific salmon life history. This method produces an estimate of the effective number of breeders in 1 year ( $N_b$ ), which, for the Pacific salmon life history, is related to generational effective size by  $N_e = gN_b$ , where  $g$  is the generation length (Waples 2002a, 2006).

A second approach adopted by some authors is to use standard methods with iteroparous species but interpret the result as  $N_b$  rather than  $N_e$ . If the sample is from a single cohort, the interpretation of  $N_b$  is relatively straightforward as the effective number of parents that produced the sample(s). However, translating  $N_b$  into an estimate of  $N_e$  is not as straightforward for iteroparous species as it is for salmon, where semelparity ensures that there is no overlap between the sets of parents each year. Generational  $N_e$  depends on lifetime variance in reproductive success, so for iteroparous species,  $N_e$  will not in general be a simple function of  $N_b$ . Furthermore, if only mixed-age samples are available,  $N_b$  refers to a potentially complex mix of overlapping sets of parents in different years, and this makes it even more challenging to relate  $\hat{N}_b$  to  $N_e$ .

A final option, adopted by many researchers, is to simply ignore the problem by either (i) not mentioning or acknowledging potential biases, or (ii) assuming that any biases will be small compared to other sources of uncertainty inherent in estimating  $N_e$ . Waples & Yokota (2007) evaluated robustness of the latter option for the temporal method, using simulated data for overlapping-generation

species with a wide range of life history types. They found that biases associated with age structure can indeed be relatively modest in some cases, provided samples are spaced enough generations (at least 3–5 or more) apart for the collective signal from drift to be large relative to the sampling biases. However, for many applications, sampling this many generations apart will not be feasible, and with a short elapsed time between samples, both the direction and magnitude of bias in  $\hat{N}_e$  can vary in a complex way with life history parameters and age composition of the samples. Palstra & Ruzzante (2008) found that published  $\hat{N}_e$  values for the temporal method were consistently higher for studies that explicitly accounted for age structure effects, suggesting that failing to account for age structure tends to downwardly bias temporal estimates of effective size. Unfortunately, comparable evaluations of sensitivity to age structure effects have not been performed for any of the single-sample estimators. Waples & Do (2010) speculated that if a random, mixed-age sample includes a number of consecutive age classes approximately equal to a generation length, a single-sample estimator should produce an estimate approximately comparable to  $N_e$  per generation. However, this conjecture remains to be evaluated quantitatively. That should be possible with a new software program (AGENE, available on request), which can calculate  $N_e$ ,  $N_b$ , and  $N$  for age-structured species based on age-specific vital rates.

*Time frame of inference.* Unlike long-term estimators, estimates of contemporary  $N_e$  apply to a specific generation or generations, and careful attention to this point is important, particularly if one is interested in comparing  $N_e$  to  $N$ . In general, single-sample estimates relate to inbreeding effective size and provide information about effective size in the parental generation; temporal estimates relate to variance effective size and provide information about the harmonic mean  $N_e$  in the interval between samples (for more discussion and details, see Waples 2005). As noted previously, interpretation becomes more complex for mixed-age samples from iteroparous species.

The LD method is an exception among the contemporary estimators in that it can potentially provide information about effective size over a wide range of time periods. LD decays by 50% per generation with recombination, so applications that use unlinked loci produce estimates that are primarily sensitive to  $N_e$  in the parental generation (Waples 2005). With physical linkage, however, LD decays much more slowly, and for tightly linked markers contemporary samples can provide insights into effective size in the distant past, provided the recombination rate is known (see Tenesa *et al.* 2007 for an example). Recent, severe bottlenecks might distort or blur this historical signal, however, and this point merits further study.

## Discussion

Demographic, evolutionary, and physical processes stratify natural populations into different classes of individuals, and these spatial-temporal stratifications can have a profound influence on effective population size. Most standard population genetic models do not fully account for the consequences of these stratifications, which has motivated the topics discussed in this document. Both local and metapopulation  $N_e$  can be important for applied conservation and management, as well as for the study of evolution. At either extreme of the isolation-connectivity continuum, there is agreement between theory and empirical results: at the weak migration limit (migration rare compared to mutation), local processes dominate, while global processes become more important as migration rates approach those for panmixia. However, many (perhaps most) natural populations are connected by levels of gene flow that are intermediate to these extremes; in this range, behaviour is not reliably predicted by theory and the relative influences of local vs metapopulation  $N_e$  depend on a variety of factors, including the time frame of interest. More research is needed to help disentangle the effects of selection and drift in metapopulations connected by intermediate levels of migration.

Levels of genetic diversity in natural populations primarily reflect long-term processes in which a balance is achieved between generation of diversity by new mutations and loss of diversity by drift. Unless isolation is virtually complete, loss of diversity through drift is determined primarily by global (or metapopulation)  $N_e$  rather than local  $N_e$ . For estimates of long-term effective size, obtaining a reliable estimate of mutation rate is the most important consideration; unless this is accomplished, estimates can be biased by orders of magnitude. Unrecognized spatial structure, or a mismatch between the geographic scale of sampling and the scale of inference, can also lead to biases, the most serious being attempting to estimate long-term local  $N_e$  in a system that experiences even modest amounts of migration.

Preliminary information indicates that at least some estimators of contemporary  $N_e$  are surprisingly robust to relatively high (approximately 10%) levels of migration; therefore, under many realistic scenarios, they might yield reliable estimates of local  $N_e$  and hence potentially important insights into local demographic and evolutionary processes. However, users must be careful not to lose track of the evolutionary processes that estimators are sensitive to. Estimators of contemporary  $N_e$  are primarily sensitive to rates of evolutionary processes occurring on the scale of one or a few generations. This means, for example, that a low estimate of contemporary  $N_e$  does not necessarily mean a local population is at risk of losing genetic diversity; in general, it will not be as long as it is connected to numerous other populations

by even low levels of gene flow. However, a population that historically was connected by migration but has recently become fragmented and isolated might be at risk of losing diversity. This emphasizes the importance of linking estimates of  $N_e$  with other information on the biology and life history of the focal species.

In most cases, stratifications within a population because of age structure probably have little effect on long-term estimators of  $N_e$ , but these effects can have a strong influence on contemporary estimators. In particular, it is very difficult to devise a scheme for taking a random sample from an entire generation in age-structured populations. More work is needed to better understand how effective size estimates for age-structured populations relate to  $N_b$  per year and  $N_e$  per generation, and how these latter two quantities relate to each other.

Additional research is also sorely needed to evaluate robustness of recent likelihood/coalescent methods for inferring historical demography and  $N_e$ . Collectively, these methods hold considerable promise for relaxing equilibrium assumptions and potentially allowing detailed insights into historical processes within and among populations over considerable time periods. However, rigorous performance evaluations and sensitivity analyses of these methods often have not been performed, no doubt due in part to enormous demands on computational power, where a single run of one data set can take days or weeks to perform. We are therefore in an exciting but challenging time where new programs are being developed faster than existing ones can be evaluated properly, and this argues for careful attention to explicit and implicit assumptions, particularly those that ignore natural process variation.

## Acknowledgements

I thank Séverine Vuilleumier, who made many useful suggestions and generously shared unpublished results, as well as two Italians from Argentina (Oscar Gaggiotti and Daniel Ruzzante) and three anonymous reviewers, who provided constructive comments on an earlier draft. This work also benefitted from discussions within the Genetic Monitoring (GeM) Working Group jointly supported by the National Evolutionary Synthesis Center (Durham, NC) and the National Center for Ecological Analysis and Synthesis (Santa Barbara, CA, USA).

## References

Avice JC, Ball RM, Arnold J (1988) Current versus historical population sizes in vertebrate species with high gene flow: a comparison based on mitochondrial DNA lineages and inbreeding theory for neutral mutations. *Molecular Biology and Evolution*, **5**, 331–344.

Balloux F (2001) *EASYPPOP* (version 1.7): a computer program for population genetics simulations. *Journal of Heredity*, **92**, 301–302.

Balloux F (2004) Heterozygote excess in small populations and the heterozygote-excess effective population size. *Evolution*, **58**, 1891–1900.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in  $n$  subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences USA*, **98**, 4563–4568.

Berendonk TU, Spitze K, Kerfoot WC (2009) Ephemeral metapopulations show high genetic diversity at regional scales. *Ecology*, **90**, 2670–2675.

Berthier K, Charbonnel N, Glaan M, Chaval Y, J-Cosson F (2006) Migration and recovery of the genetic diversity during the increasing density phase in cyclic vole populations. *Molecular Ecology*, **15**, 2665–2676.

Biro PA, Dingemans NJ (2009) Sampling bias resulting from animal personality. *Trends in Ecology and Evolution*, **24**, 66–67.

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

Charlesworth B (1980) *Evolution in Age-Structured Populations*. Cambridge University Press, Cambridge, UK.

Crow JF (1954) Breeding structure of populations. II. Effective breeding number. In: *Statistics and Mathematics in Biology* (eds Kempthorne O, Bancroft TA, Gowen JW & Lush JL), pp. 543–556. Iowa State College Press, Ames, IA.

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

Engen S, Lande R, B-Saether E (2005) Effective size of a fluctuating age-structured population. *Genetics*, **170**, 941–954.

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

Fraser DJ, Hansen MM, Ostergaard S *et al.* (2007) Comparative estimation of effective population sizes and temporal gene flow in two contrasting population systems. *Molecular Ecology*, **16**, 3866–3889.

Gaggiotti OE, Foll M (2010) Quantifying population structure using the F model. *Molecular Ecology Resources*, **10**, 821–830.

Gaggiotti OE, Vetter RD (1999) Effect of life history strategy, environmental variability, and overexploitation on the genetic diversity of pelagic fish populations. *Canadian Journal of Fisheries and Aquatic Sciences*, **56**, 1376–1388.

Hedrick PW, Gilpin ME (1997) Genetic effective size of a metapopulation. In: *Metapopulation Biology: Ecology, Genetics, and Evolution* (eds Hanski I & Gilpin M), pp. 165–181. Academic Press, San Diego.

Henn BM, Gignoux CR, Feldman MW, Mountain JL (2009) Characterizing the time dependency of human mitochondrial DNA mutation rate estimates. *Molecular Biology and Evolution*, **26**, 217–230.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747–760.

Hill WG (1972) Effective size of populations with overlapping generations. *Theoretical Population Biology*, **3**, 278–289.

Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution*, **22**, 1561–1568.

Hoffman EA, Schueler FW, Blouin MS (2004) Effective population sizes and temporal stability of genetic structure in *Rana pipiens*, the northern leopard frog. *Evolution*, **58**, 2536–2545.

- Howell N, Smejkal CB, Mackey DA, Chinnery PF, Turnbull DM, Herrnstadt C (2003) The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *American Journal of Human Genetics*, **72**, 659–670.
- Johnson DL (1977) Inbreeding in populations with overlapping generations. *Genetics*, **87**, 581–591.
- Jorde PE, Ryman N (1995) Temporal allele frequency change and estimation of effective size in populations with overlapping generations. *Genetics*, **139**, 1077–1090.
- Kuparinen A, Tufto J, Consuegra S, Hindar K, Merila J, de LeanizCG (2009) Effective size of an Atlantic salmon (*Salmo salar* L.) metapopulation in Northern Spain. *Conservation Genetics* DOI: 10.1007/s10592-009-9945-6 (published online June 2009).
- Leberg P (2005) Genetic approaches for estimating the effective size of populations. *Journal of Wildlife Management*, **69**, 1385–1399.
- 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. *Conservation Genetics*, **11**, 355–373.
- Morrissey MB, de Kerckhove DT (2009) The maintenance of genetic variation due to asymmetric gene flow in dendritic metapopulations. *American Naturalist*, **174**, 875–889.
- Nei M (1978) Estimation of heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.
- Nunney L (1999) The effective size of a hierarchically-structured population. *Evolution*, **53**, 1–10.
- Nunney L (2002) The effective size of annual plant populations: the interaction of a seed bank with fluctuating population size in maintaining genetic variation. *American Naturalist*, **160**, 195–204.
- 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 performance? *Molecular Ecology*, **17**, 3428–3447.
- Ruegg KC, Anderson EC, Baker CS, Vant M, Jackson JA, Palumbi SR (2010) Are Antarctic minke whales unusually abundant because of 20th century whaling? *Molecular Ecology*, **19**, 281–291.
- Storz JF, Beaumont MA (2002) Testing for genetic evidence of population expansion and contraction: an empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. *Evolution*, **56**, 154–166.
- Tenesa A, Navarro P, Hayes BJ *et al.* (2007) Recent human effective population size estimated from linkage disequilibrium. *Genome Research*, **17**, 520–526.
- Tufto J, Hindar K (2003) Effective size in management and conservation of subdivided populations. *Journal of Theoretical Biology*, **222**, 273–281.
- Vitalis R, Glémin S, Olivieri I (2004) When genes go to sleep: the population genetic consequences of seed dormancy and monocarpic perenniality. *American Naturalist*, **163**, 295–311.
- Vuilleumier S, Yearsley JM, Perrin N (2008) The fixation of locally beneficial alleles in a metapopulation. *Genetics*, **178**, 467–475.
- Vuilleumier S, Goudet J, Perrin N (in revision) Evolution in heterogeneous populations: from migration models to fixation probabilities.
- Wade MJ (1975) *An Experimental Study of Group Selection*. PhD Dissertation, Univ. Chicago, Chicago, IL.
- Wakeley J, Sargsyan O (2009) Extensions of the coalescent effective population size. *Genetics*, **181**, 341–345.
- Wang J (2005) Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society, Series B*, **360**, 1395–1409.
- Wang J, Caballero A (1999) Developments in predicting the effective size of subdivided populations. *Heredity*, **82**, 212–226.
- Wang J, Whitlock MC (2003) Estimating effective population size and migration rates from genetic samples over space and time. *Genetics*, **163**, 429–446.
- Waples RS (1990) Conservation genetics of Pacific salmon. III. Estimating effective population size. *Journal of Heredity*, **81**, 277–289.
- Waples RS (2002a) Effective size of fluctuating salmon populations. *Genetics*, **161**, 782–791.
- Waples RS (2002b) Definition and estimation of effective population size in the conservation of endangered species. In: *Population Viability Analysis* (eds Beissinger SR & McCullough DR), pp. 147–168. University of Chicago Press, Chicago, IL.
- Waples RS (2005) Genetic estimates of contemporary effective population size: to what time periods do the estimates apply? *Molecular Ecology*, **14**, 3335–3352.
- Waples RS (2006) Seed banks, salmon, and sleeping genes: effective population size in semelparous, age-structured species with fluctuating abundance. *American Naturalist*, **167**, 118–135.
- Waples RS, Do C (1994) Genetic risk associated with supplementation of Pacific salmonids: captive broodstock programs. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**(Suppl. 1), 310–329.
- Waples RS, Do C (2010) Linkage disequilibrium estimates of contemporary  $N_e$  using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary Applications*, **3**, 244–262.
- Waples RS, Yokota M (2007) Temporal estimates of effective population size in species with overlapping generations. *Genetics*, **175**, 219–233.
- Wares JP, Pringle PM (2008) Drift by drift: effective population size is limited by advection. *BMC Evolutionary Biology*, **8**, 235. doi:10.1186/1471-2148-8-235.
- Weir BS, Cockerham CC (1984) Estimating  $F$ -statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- Whitlock MC (1992) Temporal fluctuations in demographic parameters and the genetic variance among populations. *Evolution*, **46**, 608–615.
- Whitlock MC (2003) Fixation probability and time in subdivided populations. *Genetics*, **164**, 767–779.
- Whitlock MC (2004) Selection and drift in metapopulations. In: *Ecology, Genetics and Evolution of Metapopulations* (eds Hanski I & Gaggiotti OE), pp. 153–175. Elsevier Academic Press, London.
- Whitlock MC, Barton NH (1997) The effective size of a subdivided population. *Genetics*, **146**, 427–441.
- Whitlock MC, Gomulkiewicz R (2005) Probability of fixation in a heterogeneous environment. *Genetics*, **171**, 1407–1417.
- Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.
- Wright S (1938) Size of population and breeding structure in relation to evolution. *Science (USA)*, **87**, 430–431.