A comparison of single-sample estimators of effective population sizes from genetic marker data

Jinliang Wang

Institute of Zoology, Zoological Society of London, London NW1 4RY, United Kingdom

Left running head: J Wang

Right running head: Single-sample estimators of N_e

Key words: Effective size, drift, inbreeding, genetic markers, simulations

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

In molecular ecology and conservation genetics studies, the important parameter of effective population size (N_e) is increasingly estimated from a single sample of individuals taken at random from a population and genotyped at a number of marker loci. Several estimators are developed, based on the information of linkage disequilibrium (LD), heterozygote excess (HE), molecular coancestry (MC), and sibship frequency (SF) in marker data. The most popular is the LD estimator, because it is more accurate than HE and MC estimators, and is simpler to calculate than SF estimator. However, little is known about the accuracy of LD estimator relative to that of SF, and about the robustness of all single-sample estimators when some simplifying assumptions (e.g. random mating, no linkage, no genotyping errors) are violated. This study fills the gaps, and uses extensive simulations to compare the biases and accuracies of the 4 estimators for different population properties (e.g. bottlenecks, non-random mating, haplodiploid), marker properties (e.g. linkage, polymorphisms) and sample properties (e.g. numbers of individuals and markers), and to compare the robustness of the 4 estimators when marker data are imperfect (with allelic dropouts). Extensive simulations show that SF estimator is more accurate, has a much wider application scope (e.g. suitable to non-random mating such as selfing, haplodiploid species, dominant markers) and is more robust (e.g. to the presence of linkage and genotyping errors of markers) than the other estimators. An empirical dataset from a Yellowstone grizzly bear population was analysed to demonstrate the use of the SF estimator in practice.

Introductions

In the past few decades, genetic marker based methods have been developed and increasingly applied to estimating the effective sizes (N_e) of natural populations in widely different spatial and time scales (Schwartz *et al.* 1999; Wang 2005; Luikart *et al.* 2010; Wang *et al.* 2015). For the contemporary N_e of a local population, which is the most relevant for conservation (Luikart *et al.* 2010), the most widely applied has been the so-called temporal methods which use genetic data in two or more temporally spaced samples taken from the same population (Nei & Tajima 1981; Waples 1989; Wang 2001; Anderson 2005; Palstra & Ruzzante 2008). These methods exploit the temporal changes in allele frequencies (or differentiation between temporal samples) as information to estimate the average N_e of a population during the sampling interval. They are in general more accurate and robust (to non-random mating and genetic structure, for example) than other genetic methods (Wang 2005; Luikart *et al.* 2010), but their reliance on temporally spaced samples incurs two shortcomings. First, it is expensive to collect temporal data separated by at least one generation (Nei & Tajima 1981; Waples 1989). For long-lived species which have large generation intervals and are usually more of conservation concern, one generation could mean many years. Furthermore, such species usually have overlapping generations, and a sampling interval much larger than one generation is necessary to yield accurate N_e estimates (e.g. Nei & Tajima 1981). Second, it is of limited value in genetic monitoring (Schwartz *et al.* 2007), because the signal of temporal changes in allele frequency for a population with a drastic change of N_e in one generation will become weaker over a longer sampling interval of multiple generations, and also because of the time lag in obtaining data and N_e estimates for managements. If a population fluctuates in N_e and the sampling interval is longer than the fluctuation cycle, the temporal method may not be able to detect the fluctuations at all.

In contrast, single-sample methods exploit information of linkage disequilibrium (Hill 1981), heterozygosity excess (Pudovkin *et al.* 1996), sibship frequency (Wang 2009), or coancestry (Nomura 2008) extracted from the genotype data of a single sample of individuals taken from a population at a single time point to estimate the N_e of the population at or near that time point. They largely overcome the two shortcomings of the temporal methods, and as a result, have become popular in the past few years (Waples 2006; Waples & Do 2008, 2010; Tallmon *et al.* 2008; Wang 2009). The linkage disequilibrium (LD) based N_e estimator has become especially popular due to the work of Waples and coworkers (Waples 2006; Waples & Do 2008, 2010). They showed that the original estimator by Hill (1981) is biased (see also England *et al.* 2006), and proposed an empirical correction factor to reduce or remove the

bias (Waples 2006). They implemented the improved estimator in two software packages (Waples & Do 2008; Do *et al.* 2014), and investigated its performance for isolated populations with discrete (Waples 2006) and with overlapping (Waples *et al.* 2014) generations, and for subdivided populations with immigration (Waples & England 2011). Waples and Do (2010) concluded, from analysing simulated microsatellite data, that the LD estimator is more precise than the temporal estimator, unless the latter is applied to samples separated by many generations.

Other single-sample methods are much less popular than LD estimator. The estimator based on heterozygosity excess (HE) is known to be imprecise, yielding frequently infinitely large N_e estimates for populations known to have small effective sizes (Pudovkin *et al.* 1996; Luikart & Cornuet 1999). The estimator based on molecular coancestry (MC) has similar accuracy to HE estimator (Nomura 2008). The estimator using multiple types of information from a single sample, implemented in the software OneSamp (Tallman *et al.* 2008), has not been evaluated for accuracy against other estimators by any simulation studies. Its accuracy, robustness and other properties relative to those of the other single-sample estimators are still mysterious. The estimator based on sibship frequencies (SF) estimated from a single sample of multilocus genotypes (Wang 2009) is flexible (allowing for non-random mating, immigration, diploid as well as haplodiploid species, for example) and robust (e.g. to genotyping errors). Its accuracy, as checked by simulations, is much higher than the HE estimator, and is similar to that of the temporal methods when sampling interval is not long. The estimator is, however, much more computationally intensive than other estimators. Its accuracy in comparison with that of the LD estimator is still unknown.

In this study, I use simulations to compare the accuracies of different single-sample estimators of N_e, and to examine the effects of sampling intensities (of individuals and markers), marker information contents (marker numbers, polymorphisms and linkage), population properties (true N_e, inbreeding, and population size fluctuation) and imperfect data (e.g. allelic dropouts) on the relative accuracies of these estimators. For the first time, I explored the prospect of using many linked SNPs in N_e estimation, and the possibility of estimating N_e when it is extremely large, when the species is haplodiploid, and when the species is monoecious with a substantial selfing rate. The sensitivity of the SF estimator to prior N_e assumptions is also investigated by simulations when marker information is scarce and ample. A published empirical dataset is comparatively analysed by different single-sample N_e estimators. The results of simulation and empirical data analyses are discussed in

the context of the practical applications of N_e estimation in conservation genetics and molecular ecology.

Methods

In this section, I briefly describe the concepts of effective population size, the single-sample methods for estimating N_e, the procedures and parameter combinations used in simulations, the accuracy assessment methods, and one published empirical dataset.

1. Effective population size

In the absence of all systematic forces (i.e. mutation, migration and selection) of evolution, an infinitely large population will attain and stay in an evolutionary equilibrium in which both allele frequencies and genotype frequencies remain constant over generations. A finite population, in contrast, does not attain the equilibrium. Its allele and genotype frequencies fluctuate randomly over generations as a result of sampling a finite number of gametes to form the population at each generation. A finite population size has, among others, two genetic consequences. One is the random change of allele frequencies, called genetic drift, and the other is the increase in homozygosity and the decrease in heterozygosity at each locus, called inbreeding. In the long run, drift leads to the fixation of one allele in and the loss of all the other alleles at a locus from the population, and inbreeding leads to the loss of all heterozygotes. Drift and inbreeding are two facets of the same stochastic process in a finite population. The strength (rate) of drift and inbreeding is measured by Wright's (1931) effective population size, N_e. The N_e of a population is defined as the size of a Wright-Fisher idealised population (Fisher 1930; Wright 1931) that would lose genetic variation or become inbred at the same rate as the actual population. As a reference, a Wright-Fisher idealised population is defined with a long list of simplifying assumptions, such as a constant size of monoecious diploid individuals, discrete generations, random mating (including selfing in random amount), an equal chance of contributing offspring per parent, and the absence of all systematic evolutionary forces. Many factors affecting the drift and inbreeding processes in an actual population, such as non-random mating, unequal sex ratio, haploid or polyploid inheritance, fluctuation in population size, overlapping generations, spatial structure, and unequal individual reproductive potential, are conveniently summarized into a single parameter, N_e, and thus become irrelevant in describing the inbreeding and genetic drift of the study population.

In some simple cases, the rate of random drift in allele frequencies and the rate of decrease in homozygosity (inbreeding) are the same, and a single effective size can be used to

describe both. In some complex cases, however, the two rates can be different either temporally or permanently, and a variance effective size (N_{eV}) and an inbreeding effective size (N_{eI}) were distinguished (Crow 1954; Crow & Morton 1955) to describe the drift and inbreeding processes, respectively. N_{eV} and N_{eI} depend on the offspring and parent generations, respectively, and the former is always larger (smaller) than the latter for an expanding (shrinking) population. For an isolated or incompletely subdivided population of constant size, N_{eV} and N_{eI} are identical to each other, and to other concepts such as eigenvalue and mutational effective sizes (Whitlock & Barton 1997). Most of the simulations of this study consider an isolated population of constant size, and there is no need to distinguish N_{eV} and N_{eI}. For simulations of populations changing in size, different estimators might be estimating different quantities, and these are clarified below.

2. Single-sample N_e *estimators*

The estimator based on heterozygosity excess, HE, is the simplest in concepts and computations. When N_e is small, random genetic drift is expected to lead to a difference in allele frequency between male and female parents, which in turn results in an excess of heterozygotes in the offspring compared with Hardy-Weinberg equilibrium (Robertson 1965). The extent of heterozygote excess is expected to be inversely proportional to the actual N_e of the parental population, and thus can be calculated from marker data to estimate N_e. The estimator was derived by Pudovkin *et al.* (1996), implemented in several computer programs (e.g. Zhdanova & Pudovkin, 2008; Jones & Wang 2010; Do *et al.* 2014), and evaluated in several studies (e.g. Luikart & Cornuet 1999; Wang 2009).

The estimator based on molecular coancestry, MC, was proposed by Nomura (2008). The average coancestry of an isolated population is expected to increase at a rate inversely proportional to the N_e of the population. Nomura proposed a method to quantify this rate from a single sample of multilocus genotypes, and the estimated rate was converted to an estimate of N_e. Briefly, he proposed to select putative nonsib pairs from all of the possible pairs in a sample of individuals based on their molecular coancestry. The selected nonsib pairs are then used as reference to calculate the average parent-based coancestry, $\hat{f}_1$. The inbreeding effective size is then estimated by $\widehat{N}_e = 1/(2\hat{f}_1)$.

The estimator based on linkage disequilibrium, LD, was derived by Hill (1981). In the absence of all systematic forces, the frequency of an allele in a finite population changes (drifts) randomly over time and also becomes correlated randomly with that of another allele at a different locus. This correlation is called linkage disequilibrium, and has a simple

functional relationship with the recombination rate, *c*, between loci and the effective population size (Hill 1981). When LD is calculated from a sample of multilocus genotypes, it can then be used to estimate N_e if *c* is known. In the case of unlinked markers (*c*=0), Hill's method was much improved by Waples and coworkers (Waples 2006; Waples & Do 2008 2010) to reduce biases caused by small sample sizes. The improved estimator, which assumes *c*=0, was studied by simulations for accuracy (e.g. Waples & Do 2009) and for robustness to population subdivision (Waples & England 2011) and overlapping generations (Waples *et al.* 2014). The present study compares this improved estimator, which assumes unlinked markers, with other single-sample estimators.

The estimator based on sibship frequency, SF, was derived by Wang (2009) using both the inbreeding and genetic drift approaches. In essence, the SF method for estimating N_e is the analogue to the mark-capture-recapture method for estimating population census size (Luikart *et al*. 2010). The capture unit is a pair of individuals, and the recapture rate is the frequency of a sibling dyad. Populations of a small (large) N_e will have a high (low) frequency of sibling dyads. Implemented in the software package Colony (Jones & Wang 2010), the SF estimator has several advantages, such as its ability to account for inbreeding, its flexibility for use in diploid as well as haplodiploid species, its wide application scope of codominant as well as dominant markers with or without linkage and with or without genotyping errors, and its robustness to population subdivision (Wang 2009). However, it is computationally demanding, especially for a large sample under the polygamous mating system, because a simulated annealing algorithm is employed to assign the sampled individuals into full sibships nested within half sibships from the multilocus genotype data (Wang & Santure 2009).

The SF estimator requires ample marker information to obtain unbiased and precise N_e estimates (Wang 2009). When marker information is scarce (say, less than 10 microsatellites) and the actual N_e is large (such that the actual siblings are rare in a sample), it tends to underestimate N_e because of the over-assignment of sibship (Wang 2009). This problem becomes severe when both males and females are polygamous such that half siblings rather than full siblings are dominant in determining N_e but are unfortunately difficult to estimate. An approach to reducing the problem is to introduce a prior sibship distribution, based on prior knowledge of the study population, to penalize sibship assignment and encourage non-sibship assignment, as detailed below.

3. A prior for SF estimator

The accuracy of the SF based N_e estimator depends critically on the qualities of sibship assignments. Sibship over- and under-assignments lead to a higher and a lower sibship frequency and thus an under- and over-estimate of N_e respectively. Both type I (false sibship assignments) and II (false non-sibship assignments) errors are expected to diminish with an increasing amount of marker information, as verified by simulations (Wang & Santure 2009). However, when marker information is scarce and sample size is large, both types of assignment errors could occur. In a sample taken at random from a large population, a sibship assignment analysis usually makes much more type I errors than type II errors. In such a sample, most individuals are unrelated or only remotely related (such as cousins) when the current sample (population) is used as reference (Wang 2014). However, they may happen to display similar genotypes compatible with the sib relationship when marker information is insufficient, and may thus be erroneously assigned to be sibs. To reduce these errors which lead to an underestimated N_e, I use Ewen's (1972) sampling formula

$$Prb[\boldsymbol{a}_n] = \frac{n!}{\theta_{(n)}} \prod_{j=1}^{n} \left(\frac{\theta}{j}\right)^{a_j} \frac{1}{a_j!}, \quad (1)$$

as a prior to model the distribution of offspring among male (or female) parents contributing to the sample. In (1), θ is the sole parameter of the distribution, n is sample size (number of sampled offspring), a_j (j=1, 2, …, n) in $\mathbf{a}_n$=(a_1, a_2, …, a_n) is the number of parents of a given sex who each has exactly j offspring (thus $\sum_{j=1}^{n} j a_j \equiv n$) in the sample, and $\theta_{(n)} = \prod_{i=0}^{n-1}(\theta + i)$.

In (1), the number of parents contributing to the sampled n offspring is expected to be $K_n = \prod_{j=1}^{n} a_j$. Obviously, $1 \leq K_n \leq n$, and the average number of offspring per contributing parent is n/K_n. For a given sample of n offspring, θ determines the offspring distribution among parents, (1), entirely. When $\theta = 0$, we have $K_n = 1$, and all of the n sampled offspring come from a single parent with probability 1. When $\theta \to \infty$, we have $K_n = n$, and the n offspring come from n distinct parents (i.e. no siblings) with probability 1. When $\theta = 1$, distribution (1) is precisely that of the integer partition induced by a uniformly distributed random permutation.

For a sample of offspring taken at random from a population, K_n is intuitively expected to increase with increasing values of n and N_e. After extensive experimentation, I arrived at the empirical function

$$\widehat{K}_n = n - 70\left(1 - e^{-1/N_e}\right) + 0.2(1 - e^{-0.01/r}) \quad (2)$$

to determine $\widehat{K}_n$ from *n* and N_e, where N_e is the prior effective size and $r = n/N_e$ is the ratio of sample size to prior effective size. When both *n* and N_e are very small, $\widehat{K}_n$ as calculated above can be smaller than 1. In such a rare case, it is reset to the minimum value 1. Given $\widehat{K}_n$, the maximum likelihood estimate of θ is obtained by solving the equation $\prod_{i=0}^{n-1} \hat{\theta}/(\hat{\theta} + i) = \widehat{K}_n$, which can then be used in calculating prior (1).

I attach a weight, *W*, to prior (1) for controlling its importance relative to data in a sibship analysis. The idea is to reduce the reliance on the prior with an increase in sample size relative to N_e, $r=n/N_e$. The rationale is that, with an increase in *r*, the frequency of type I errors relative to type II errors decreases, and the prior intended to reduce type I errors becomes less important. After some experimentation, I arrived at the empirical weight function,

$$W = 0.5 - 0.15Ln(r) + 0.2e^{-10r} - 10e^{-5/r} + 0.01(1 - e^{-r})Ln(R), \quad (3)$$

where *R* is the sex ratio (defined as the ratio of the numbers of breeders of the minority sex and majority sex, $R \leq 1$). *W* decreases fast with an increasing *r*, but very slowly with an increasing *R* (Figure 1). When *n* is not much larger than prior N_e, $W>0$, and a prior of $(Prb[\boldsymbol{a}_n])^W$ is used in sibship assignment analysis. Otherwise, $W \leq 0$, and no prior is used in sibship analysis.

When prior information about N_e and *R* is unavailable, default priors are also provided in the Colony program to reduce type I errors. These priors (Wang & Santure 2009) use the same formula (1), with parameter θ being calculated from the assumption of $K_n=n$ (or alternatively, the average sibship size is 1 individual) or a user provided estimate of average sibship size.

4. The SF estimator for monoecious species with selfing

No single-sample N_e estimators are available for monoecious species with selfing occurring at a substantial rate. Following the approach of Wang (2009), however, an equation of N_e in terms of full and half sibship frequencies can be derived for monoecious species with selfing,

$$N_e = \frac{4}{(1+\alpha)(Q_1+Q_2+2Q_3)}, \quad (4)$$

where α is the parameter for the deviation from Hardy-Weinberg equilibrium in genotype frequencies (equivalent to Wright's (1969) F_{IS} statistic) due to selfing, and Q_1, Q_2 and Q_3 are the frequencies of paternal half-sib, maternal half-sib and full-sib dyad frequencies. Note the relevant quantity is the total frequency of two individuals sharing the same parent, $Q_1 + Q_2 + 2Q_3$, not the separate values of Q_1, Q_2 and Q_3. Therefore, it is unnecessary to distinguish maternal and paternal relationships in estimating N_e. In fact, it is difficult to differentiate paternal from maternal half sibs using autosomal marker data only. The likelihood method for sibship analysis (Wang & Santure 2009) assigns sibship, leading to an estimate of $Q_1 + Q_2 + 2Q_3$. It also infers the selfing/outbreeding status of each individual and thus yields a direct estimate of selfing rate, $\hat{s}$ (Wang *et al*. 2012). Given $\hat{s}$, α is estimated as $\hat{\alpha} = \hat{s}/(2 - \hat{s})$ (Haldane 1924; Caballero 1994). In the special case of no selfing (s=0), α =0 and (4) reduces to the formula for a dioecious species with equal numbers of males and females (eqn 10 in Wang 2009, see also Waples & Waples 2011).

Estimator (4) performs well, providing nearly unbiased and accurate estimates of N_e, when selfing rate s is not high. When s is high, however, it underestimates N_e substantially. This is understood by considering the chance that non-sibs have identical multilocus genotypes and are thus falsely assigned sibship. This chance increases rapidly with an increasing selfing rate s. Given s, the probability that an offspring comes from a lineage that has undergone g consecutive generations (counting backward in time) of self-reproduction before an outcrossing event is $s^g(1 - s)$. For a high selfing rate, s=0.95 for example, the probability that $g \geq 10$ is 0.6, and the inbreeding coefficient of the grand parent is at least $1 - 0.5^8$=0.996. This means the grand parent is highly likely to be homozygous across all marker loci. When s=0.95, therefore, any two offspring coming from different parents by self-reproduction (thus non-sibs) but from the same grandparent display an identical multilocus genotype and thus are mis-assigned full sibship with a probability of ≥0.6. This numerical example provides an underestimate of type I errors, because two non-sibs with non-identical multilocus genotypes, if similar enough due to selfing, can still be mis-assigned as full or half sibs.

To correct for the over-occurrence of type I errors and the underestimation of N_e when s is high, I took an empirical approach by deriving a correction factor C, which is multiplied by the original N_e estimate (4) (assuming α=0) to give the final estimate. I generated simulated data for a monoecious population under different s values in the range [0, 0.995],

obtained $\hat{s}$ and $\widehat{N}_e$ from (4) (assuming α=0) by sibship analysis of simulated data, and fitted the theoretical N_e calculated by eqn (7) below to $\widehat{N}_e$ as a function of $\hat{s}$. This analysis (data not shown) yielded that $C = 1/(1+\hat{\alpha})$, $C = 1/(1.05 - 2b + 6b^2 - 5b^3 - 15b^4)$ where b=$\hat{s} - 0.5$, and $C = 1/(0.5 - b - 35b^2 + 300b^3 - 200b^4)$ where b=$\hat{s} - 0.9$, when $\hat{s} = [0,0.5)$, $\hat{s} = [0.5, 0.9)$, and $\hat{s} = [0.9,1)$, respectively. N_e is first estimated by (4) using estimated sibship frequencies and assuming α=0, and then the estimate is multiplied by the correction factor chosen according to $\hat{s}$ to give the final N_e estimate.

5. Simulation procedures

Three species models, dioecious diploid (DD), dioecious haplodiploid (DH), and monoecious diploid (MD), were considered in simulations. For DH model, females and males were assumed to be diploid and haploid respectively. For MD model, self-reproduction and outbreeding occurred at frequencies *s* and 1-*s* respectively. The population was composed of $N_1(t)$ males and $N_2(t)$ females for dioecious species and of $N(t)$ individuals for monoecious species at a discrete generation *t*. Population size and sex ratio (for dioecious only) were assumed constant in most simulated scenarios, but a bottleneck in the parental population was also considered (below). Individuals in the founder generation (t=0) were assumed to be unrelated and non-inbred, and the genotype of each individual was generated for *L* loci. Each locus *l* (l=1~*L*) was assume to have *A* codominant alleles with frequencies in a uniform Dirichlet distribution. For microsatellites, I assumed *L* was small, A was large, and the loci were unlinked. For SNPs, I assumed *L* was large, *A*=2, and the loci were equally spaced in a genome of genetic map length *M* Morgans. Assuming Hardy-Weinberg and linkage equilibrium, the multilocus genotype of each founder individual was generated by drawing a paternal and a maternal gene at each locus independently from the given allele frequency distributions.

Starting from the founder generation, a number of g=10 (or g=1000 in the case of linked SNPs) generations were simulated before a sample of offspring was taken at random from the population at generation g+1 for genotype analysis and N_e estimation. More generations (larger *g* values) before sampling are not necessary, and do not change the results essentially because none of the 4 estimators requires an equilibrium between drift and mutation. For the simulated unlinked and linked markers, g=10 and g=1000 are sufficient to attain stable values of LD. At each generation *t* (=1~*g*), each of the $N_i(t)$ individuals for the dioecious case (i=1,2 for males and females) or the $N(t)$ individuals for the monoecious case was generated independently. To generate an offspring in the dioecious case, the father and

mother were drawn at random from the $N_1(t\text{-}1)$ males and $N_2(t\text{-}1)$ females respectively. The pedigree of the offspring was recorded to calculate coancestry and simulated N_e (below), and the multilocus genotype of the offspring was generated from those of its father and mother following Mendelian segregation law. For linked SNPs, the number of crossovers in generating a gamete was drawn from a Poisson distribution by assuming Haldane's (1919) map function, with the mean number of crossovers being *M*. The locations where crossovers occurred were randomly chosen along the chromosome. In the monoecious case, an offspring was determined to come from self-reproduction and outbreeding at frequencies *s* and 1-*s* respectively. In the former case, a single parent was drawn at random from the $N(t\text{-}1)$ individuals to produce both a male and a female gametes which united to produce the offspring. In the latter case, the same procedure as in the dioecious case was followed, except both male and female parents were drawn at random without replacement (i.e. selfing excluded) from the same $N(t\text{-}1)$ individuals. At generation g+1, a number of *n* diploid individuals were generated and their multilocus genotypes were used by various N_e estimators to calculate N_e.

The simulation program was checked in several ways to ensure it worked properly. First, the average coancestry at generation *t* in the DD model was calculated from simulated pedigrees by $\bar{G}_t = \frac{1}{4}\sum_{x=1}^{2}\frac{1}{N_x(t)}\sum_{y=1}^{2}\frac{1}{N_y(t)}\sum_{i=1}^{N_x(t)}\sum_{j=1}^{N_y(t)} G_{ix,jy}(t)$, where $G_{ix,jy}(t)$ is the coancestry between individuals *i* of sex *x* and *j* of sex *y* (*x*,*y*=1,2 for males and females). The effective size of generation *t* is calculated by $N_e(t) = 0.5(1-\bar{G}_t)/(\bar{G}_{t+1}-\bar{G}_t)$. For constant population size, this calculated N_e is expected to be constant, equal to the theoretical value (below). For DH and MD models and for the bottleneck model, the coancestry based N_e was calculated similarly. Second, the F_{ST} between generations 0 and *t* is expected to be $1-(1-1/(2N_e))^t$ for an isolated population with effective size N_e. This prediction was checked against the value calculated from the multilocus genotype data. Third, N_e was also calculated from the known (simulated) sibship frequencies and compared with the theoretical value (below). These checks verified that the simulation program behaved normally as expected.

6. Simulation parameter combinations

Many factors could potentially affect the performance of a N_e estimator. These include population properties (e.g. true N_e, mating system, sex ratio, distribution of family sizes, inbreeding, genetic structure), sample properties (e.g. numbers of sampled markers and

individuals), marker properties (e.g. polymorphisms and linkage), genotyping problems (e.g. allelic dropouts and missing data), as well as parameter settings of each N_e estimators (e.g. P_{crit} for HE and LD estimators, priors for SF estimator). The number of combinations of these factors is combinatorically large to explore. Therefore, I chose to investigate the impact of only one or two factors in a set of simulations, keeping other factors constant. The parameter values used in different sets of simulations, detailed below, are summarised in Table 1.

Simulation 1, number of loci: It was intended to compare the accuracy of different methods when different numbers (L) of microsatellite markers were used in N_e estimation. It was also used to investigate if a method was statistically consistent, being increasingly more accurate with an increasing amount of information. Other parameters in the simulation are typical of current N_e estimation practices (Table 1).

Simulation 2, sample size: Together with L, sample size n determines the sampling intensity and effort, and is an important determinant of information content. This simulation compares the qualities of N_e estimates by different methods applied to samples of different numbers of individuals. It considered a relatively large population (N_e =500), and correspondingly used a set of highly informative markers (L=20, A=10, Table 1).

Simulation 3, actual N_e: This simulation was used to investigate whether single-sample methods could be used to obtain reasonably good N_e estimates for populations of medium to large sizes. Sample size and number of loci were fixed at modest values of n=100 individuals and L=20, respectively, and other parameters (Table 1) were also fixed at values typical of practice. The actual N_e varied hugely in the range [10, 31250].

Simulation 4, prior N_e: The optimal prior for the SF estimator requires prior values of N_e and sex ratio R to calculate the prior probability (by (1-3)) of a sibship configuration. In reality, N_e and R are unknown, although estimable by methods such as LD. Fig 1 shows that the prior for sibship assignments is little affected by R (as was also checked by simulations, data not shown), but is strongly influenced by prior N_e. This set of simulations was used to investigate the sensitivity of SF estimator to prior N_e values. Simulated data generated with fixed parameters of N_e =50, L=10 and L=50 (other parameters are in Table 1) were analysed by SF method assuming different prior N_e values (5, 10, 20, 40, 80, 160). The data were also analysed by SF method, using N_e estimates from the LD method as priors and using the default priors.

Simulation 5, population fluctuations: SF, MC and HE methods estimate the parental N_e, while LD method estimates an average N_e of the population in the past few generations (Wang 2005). The number of generations relevant for LD based N_e estimates depends on the

linkage of markers, and how much (i.e. weight) the genetic drift (N_e) of each generation contributes to the estimated average N_e is unclear. This set of simulations considered a change in the parental population size to investigate how different estimators respond to the change. A population was assumed to have constant and equal numbers of male and female breeders of N_1=N_2=50 (thus N_e=100), except in the parental generation in which N_i (i=1,2) and thus parental N_e were changed by G%. The other parameters are in Table 1. The theoretical N_e was calculated and presented for the parental population.

Simulation 6, linked SNPs: This simulation considered the use of many SNPs, and the effect of linkage on the estimators. A fixed number of 1000 SNPs were assumed equally spaced in a genome of various map lengths, in the range [1, 64] Morgans. The other parameters are in Table 1. The population was simulated for a number of g=1000 generations before sampling, and a high mutation rate of u=0.01 was assumed at each generation to maintain the polymorphisms of SNP loci.

Simulation 7, monoecious species with selfing: This simulation considered a monoecious species (MD model) with different selfing rates, in the range [0, 1]. The other parameters are in Table 1. The simulation was used to investigate the accuracies of different estimators in the face of non-random mating (selfing).

Simulation 8, haplodiploid species: This simulation considered a species which had haploid males and diploid females (DH model). A sample of 50 diploid offspring were sampled and genotyped at a variable number of marker loci, L=[5,40], for N_e estimation, with the other parameters fixed at constant values (Table 1).

Simulation 9, allelic dropouts: This simulation considered the robustness of different estimators to allelic dropouts, a common problem for microsatellites (Bonin *et al*. 2004), especially with noninvasive samples (e.g. faeces, hair). For each sampled individual at each locus, allelic dropouts were assumed to occur at a rate D (=0, 0.025, 0.05, 0.1, 0.2, 0.4) during PCR. Under the allelic dropout model, a heterozygous genotype, AB, was observed (genotyped) to display a phenotype AB, AA and BB at probabilities 1-2d, d and d (where d=D/(1+D)), respectively (Wang 2004), when double dropouts were ignored. Double dropouts (where both alleles at a single-locus genotype dropout) rarely occur and, if they do, can be easily detected and thus rectified by regenotyping in practice. A homozygous genotype, say AA, was not affected by allelic dropouts. This allelic dropout model was applied to the genotype at each locus of each individual independently. The other simulation parameters were fixed at constant values (Table 1).

7. Theoretical N_e values

For the case of constant population size, the theoretical effective size is predicted to be (Caballero 1994)

$$N_e = \frac{4N_1N_2}{N_1+N_2} \quad (5)$$

$$N_e = \frac{9N_1N_2}{4N_1+2N_2} \quad (6)$$

$$N_e = \frac{N}{1+s/(2-s)} \quad (7)$$

for the DD, DH and MD models considered in the simulations described above, where *s* is the simulated selfing rate. In the case of a bottleneck (population fluctuation), the inbreeding effective size of the parental generation can still be calculated by (5-7), where N_i (i=1,2) or N refers to the size of the parental population.

8. Accuracy assessments

The simulated data were analysed by HE, MC, LD and SF methods. The first 3 methods were implemented in the software package NeEstimator (Do *et al.* 2014), which was used in analysing the simulated and empirical data of the present study. To reduce bias and increase accuracy, LD and HE estimators were calculated by NeEstimator by using different default values of P_{crit} (0.1, 0.05, 0.02, and 0.01) to screen out rare alleles. It is unclear which P_{crit} value is the best. It is possible that the best P_{crit} value varies, depending on sample sizes, number of loci and allele frequency distributions at each locus. In the present simulation study, the P_{crit} value that yielded the most accurate (in terms of RMSE, below) estimates was adopted and the corresponding results were reported. Negative estimates from MC, LD and HE estimators were taken as infinitely large N_e values. For the SF method, the simulated N_e value was used in calculating prior (1) and its weight (3), except when explicitly stated. The software package Colony (Jones & Wang 2010) was used to analyse the simulated and empirical data for sibship assignments and N_e estimates.

The quality of a N_e estimator can be measured by its bias, *B*, and variance, *V*, of $1/\widehat{N}_e$. The quality statistics are calculated from $1/\widehat{N}_e$ rather than $\widehat{N}_e$ because the latter can be infinitely large (and can be in a bimodal distribution) and, more importantly, because the former is more relevant in most applications (Wang & Whitlock 2003). *B* and *V* are calculated by

$$B = \frac{1}{\hat{N}_e^H} - \frac{1}{N_e},$$

$V = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{\widehat{N}_e^H} - \frac{1}{\widehat{N}_{ei}}\right)^2$,

where $\widehat{N}_{ei}$ is the estimated effective size in the *i*th (*i*=1~ *m*) replicate, N_e is the theoretical (simulated) effective size calculated by (5-7), and $\widehat{N}_e^H$ is the harmonic mean of $\widehat{N}_{ei}$ calculated by $\widehat{N}_e^H = \left(\frac{1}{m}\sum_{i=1}^{m}\frac{1}{\widehat{N}_{ei}}\right)^{-1}$. From hereafter "mean" N_e estimate always refers to this harmonic mean estimate, $\widehat{N}_e^H$.

The overall accuracy of an estimator is measured by root mean squared error, $\text{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{\widehat{N}_{ei}} - \frac{1}{N_e}\right)^2}$.

It includes both the bias and the variance, as RMSE=$\sqrt{B^2 + V}$, and measures how far the estimates $1/\widehat{N}_{ei}$ differ from the true parameter value $1/N_e$. From hereafter the word "accuracy" signifies the level of measurement that yields true (no systematic errors, *B*=0) and consistent (no random errors, *V*=0) results, quantified by RMSE. In this study, $\widehat{N}_e^H$ and RMSE are reported for each simulated parameter combination with *m*=100 replicates.

9. Yellowstone grizzly bears

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, and used the data to study the population demographic trajectories. Herein I analyse the genotype data of the sampled individuals born in the periods 1988-1990, 1998-2000, and 2008-2010. The 3 corresponding samples contain 46, 92 and 59 individuals, respectively. Because the population has overlapping generations and each sampling period has only 3 years, the estimates are effective numbers of breeders, N_b, rather than N_e. For simplicity, however, I still call the estimates N_e hereafter.

The analyses by LD, HE and MC estimators are straightforward. Estimation by SF requires, however, a prior N_e, which is unfortunately unknown. It is possible to use the LD estimate as the prior, as I did for some simulations. To demonstrate the usefulness of SF when no prior N_e is available or the LD estimate is deemed unreliable (e.g. negative or infinitely large N_e estimates for a population known to be small) as a prior, I analysed each of the 3 samples by assuming widely different prior N_e values. The plots of N_e estimates by SF estimator as a function of the prior values reveal the most likely N_e. I also analysed each of the 3 samples by using the default prior without the need of prior N_e values.

The sample sizes are highly variable among the 3 periods. To investigate whether N_e estimates are affected by sample size, I generated 100 bootstrapping samples, each of 50 individuals, from the original sample of 92 individuals for the period 1998-2000. The samples were analysed by SF estimator using different prior N_e values. For each prior, the harmonic mean of the 100 estimates was reported. Harmonic mean N_e estimates from the other three methods were also obtained and reported.

Results

Number of loci

Both MC and HE underestimate N_e substantially, even when many (L=40) highly polymorphic (A=10) markers are used in the estimation (Fig 2). While HE becomes less biased slowly with an increasing L, MC always underestimates N_e by about 60% irrespective of L. In contrast, LD and SF overestimate N_e slightly when L=5, and become essentially unbiased when $L \geq 10$.

MC and HE methods are also highly imprecise, yielding infinite N_e estimates frequently even when marker information is ample (L=40 and A=10). As a result of the low precision and high bias, MC and HE methods are much less accurate than LD and SF methods. Their RMSE values are an order higher than those of LD and SF methods (Fig 2). SF is more accurate than LD by several folds when L is small, but the difference decreases with an increasing L. The maximum advantage of SF over LD in accuracy occurs at L=10, probably due to the contribution of the prior in SF (see Fig 5 below).

Both sample size n (relative to the actual N_e, see below) and number of loci L affect the bias of MC and HE methods. Further simulations showed that, at L=10 and other parameter values as in Fig. 2, the mean HE estimates are 88 and 97 and the mean MC estimates are 51 and 52 when n=200 and 1000, respectively. While HE becomes essentially unbiased, MC still underestimates N_e substantially, when a large sample relative to effective size (i.e. large n/N_e ratio) is used.

Sample size

LD overestimates and SF underestimates N_e when sample size n is much smaller than the actual N_e (Fig 3). The biases of both methods reduce quickly with an increasing n. Both MC and HE methods underestimate N_e greatly, and only the latter shows a bias decreasing with an increasing n. In the entire range of simulated sample sizes, SF and LD methods are more accurate than MC and HE methods by roughly one order (Fig. 3).

Actual N_e

When n is not much smaller than N_e (i.e. n/N_e not much smaller than 1), LD, SF, and HE estimators are unbiased but MC estimator is downwardly biased. HE method and LD and SF methods start to severely underestimate N_e when $n/N_e \leq 1$ and $n/N_e \leq 0.015$ respectively (Fig 4). It seems impossible for HE to provide unbiased estimate of N_e when the sample size is equal to or smaller than the actual N_e. However, both LD and SF are nearly unbiased, except when the actual N_e is many times larger than n. In such a case, increasing marker information (L and A) helps to reduce bias. For LD which could yield negative estimates of N_e, the bias could also be reduced by not converting negative estimates to infinity. However, this treatment would inevitably decrease the precision, and thus the overall accuracy (RMSE) of the LD estimator.

SF estimator gives the most accurate estimates in the entire range of the actual N_e, [10, 31250]. It is always at least one order more accurate than the HE and MC methods. LD is only slightly more accurate than HE method when the actual N_e is very small. However, its accuracy advantage over HE increases rapidly with an increasing N_e.

Prior N_e

When markers are not highly informative (L=10), the SF estimator depends on the assumed prior N_e value (Fig 5A, B). It increases with an increasing prior N_e. It is smaller than the actual N_e =50 but larger than the prior N_e when the latter is much smaller than the former, and *vice versa* (Fig 5A). Over the large range of prior N_e values [5, 160], however, SF is still less biased than MC and HE, and provides more accurate estimates than MC and HE by roughly one order. Compared with LD, SF underestimates and overestimates N_e substantially when prior N_e is much smaller and larger than the actual N_e, respectively. In both cases, SF is less accurate than LD (Fig 5A, B). However, when the prior N_e is not much different from the actual N_e, or when LD estimates are used as the prior, or the default priors (which does not require prior N_e values) are used, SF estimator becomes more accurate than LD.

The default prior gives much better N_e estimates than the prior that uses LD estimates as prior N_e values. It leads to less accurate N_e estimates than the prior with assumed prior N_e values only when the assumed prior N_e values are close to the actual N_e values.

When markers are highly informative (L=50), the SF estimator is almost unaffected by the priors (Fig 5C, D). The SF estimator is always unbiased and much more accurate than the other estimators, no matter it uses the default prior, the prior with assumed widely different N_e values, or the prior with LD estimates as prior N_e values.

Simulations presented in Fig 5 were conducted with a small sample size of n=50. The SF estimator becomes increasingly independent on the prior with an increasing sample size

relative to the actual N_e (i.e. n/N_e ratio). This is understandable from Fig 1, which plots the weight of the prior used in likelihood sibship assignments as a function of n/N_e ratio. When $n/N_e \geq 1.59$, the weight becomes zero and no prior is actually used in sibship assignments and thus SF estimator is completely independent of prior N_e values.

Several other simulations using different values of n, N_e, L, and A yielded results similar to those shown in Fig 5.

Population fluctuations

When marker information is relatively ample (n=100, L=20, A=10), a drastic change in the parental population size, in the range [-40%, 40%], was undetectable by MC estimator, but was detected by the other three estimators (Fig. 6). Theoretically, SF and HE estimate the N_e in the parental generation, and thus the results are not surprising. LD is determined by a number of generations of drift occurred before sampling, and thus in theory it should estimate the (weighted) average N_e in the past few generations. Despite this, however, Fig. 6 shows that the LD method does not overestimate and underestimate parental N_e much when it is greatly decreased and increased (by 40%), respectively. This is perhaps because half of the LD detectable from a sample of individuals comes from the parental generation and the other half from previous generations.

Linked SNPs

With a decreasing genome size M and thus an increasing degree of linkage among the 1000 SNPs, the LD estimator increasingly underestimates N_e and becomes inaccurate (Fig 7). The other three estimators do not use linkage disequilibrium as information and are thus much less affected by M, except when it is very small. The HE estimator is almost unbiased, irrespective of M, but has a low precision. As a result, its overall accuracy is higher than LD only when $M<2$ (Fig 7). MC estimator provides the worst estimates in the entire range of M=[1, 64]. With an increase in M (and thus a decrease in linkage), the quality of the LD estimator (which assumes unlinked markers) approaches that of SF.

Monoecious species with selfing

The LD, HE and MC estimators are not developed for application to monoecious or dioecious species with non-random mating, such as close relative mating (including selfing). When blindly applied, LD and MC underestimate and HE overestimates N_e increasingly with an increasing selfing rate, s (Fig 8). The HE estimates of N_e become infinite when $s>0.1$. As a result, these estimators have a rather low overall accuracy measured by RMSE. In contrast, the SF estimator applies to monoecious species with selfing, and applies to high selfing rate when the correction factor is used. For the entire range [0, 0.98] of s, the SF estimator only

slightly underestimates N_e and has an overall accuracy much higher than the other estimators (Fig 8).

Haplodiploid species

The current LD, HE and MC estimators assume a dioecious species with diploid males and females, or a monoecious species of diploid individuals with selfing occurring at random (i.e. at rate 1/*N*). When applied to haplodiploid species, these estimators underestimate N_e greatly, and as a result are highly inaccurate (Fig 9). In contrast, SF estimator allows for different species models, and gives N_e estimates that are only slightly smaller than the simulated (true) value and are highly accurate.

Allelic dropouts

Allelic dropouts at marker loci cause LD and HE estimators to severely overestimate N_e, but have much less effects on SF and MC estimators (Fig 10). Allelic dropouts induce an apparent deficiency of heterozygotes, and thus lead to dramatic overestimates of N_e by HE. The N_e estimates by HE become infinitely large when $D \geq 0.05$. Dropouts also result in overestimates of N_e by LD estimator. Among the 4 estimators, SF is the only one that has a built-in model to account for dropouts, and thus is robust to these genotyping errors. It yields slight overestimates of N_e only when the actual dropout rate is very high, much higher than the assumed dropout rate (0.05) used in sibship analysis. When the actual dropout rate was used, SF always gave an almost unbiased N_e estimate (data not shown).

The relative accuracies of the 4 estimators follow a pattern similar to relative biases (Fig 10). Overall, SF performs much better than other estimators in the presence of dropouts. LD estimator can tolerate allelic dropouts when they occur at a rate smaller than 0.05. HE estimator is highly vulnerable to dropouts. Although MC is insensitive to allelic dropouts, its performance is always very poor.

Yellowstone grizzly bears

The SF estimates of N_e are relatively insensitive to the assumed prior N_e values for each of the three sampling periods (Fig 11), because the 20 microsatellite markers are sufficiently informative. The estimates increase slowly with an increasing prior N_e; they are larger and smaller than prior N_e values when the latter are small and large, respectively. The best estimates are provided by the cross points where the estimates are equal to the priors (Fig 11). For periods 1988-1990, 1998-2000, and 2008-2010, the best estimates are 69, 90 and 100 respectively. The corresponding LD estimates and SF estimates using the default prior are slightly smaller, which are 48, 60 and 77 respectively for LD and 44, 85 and 59 respectively for SF with default prior. In contrast, the MC estimates are always very small (from 11 to 21),

and the HE estimates are either very large or small, with values 1178, 35 and ∝ for periods 1988-1990, 1998-2000, and 2008-2010 respectively.

The sample size for period 1998-2000 is 92, much larger than those (46, 59) of the other two periods. The harmonic mean N_e estimate over 100 bootstrapping samples (each of size 50), plotted as a function of the prior N_e assumed in the estimation, shows that the best N_e estimate from SF remains the same, about 90. The harmonic mean N_e estimates from SF with default prior, LD, HE and MC are 88, 66, 36, 11, respectively, also very close to the estimates from the original sample of 92 individuals, which are 85, 60, 35, and 11, respectively. This shows that sample size does not affect the results for this dataset.

Discussion

In this study, I propose to use Ewen's sampling formula as a prior to reduce type I errors in the likelihood based sibship assignments, and thus to reduce the overestimation of sibship frequencies and the underestimation of N_e in the difficult situation where a small sample of individuals is drawn from a large population and genotyped for a small number of loci. For the first time, the SF method was compared with LD and other single-sample methods by analysing data simulated in widely different scenarios, including a large actual N_e, many linked SNPs, genotyping errors, different species models, and the presence of bottlenecks and close inbreeding (selfing). These scenarios are realistic, but have not been studied in previous simulations of single sample estimators. The main findings of the simulation study are summarized in Table 2. An empirical dataset was analysed comparatively by different single-sample estimators. In this section, I discuss the findings of this study and implications to estimating N_e from a single sample of individuals in practice.

Prior N_e: In a small sample of *n* individuals taken at random from a large population of effective size N_e, the sibling frequencies are expected to be low and the frequencies of non-sibs, including distant relatives such as cousins, are expected to be high. When this prior information is not used and marker information is insufficient (i.e. few loci, low polymorphism, and small *n* relative to N_e), a sibship analysis would make much more type I errors (false sibs) than type II errors (false non-sibs). With a small n/N_e ratio, non-sibs are frequent, but when characterized by just a few markers of low polymorphism, they could display similar or even identical multilocus genotypes and thus could be erroneously inferred as siblings. These type I errors become more frequent with a smaller n/N_e ratio and a smaller amount of marker information.

This study uses Ewen's sampling formula as a prior to reduce type I errors, and thus to reduce the overestimation of sibship frequency and the underestimation of N_e. Both the parameter θ (eqn 2) and the weight of the prior are designated to depend on n/N_e (eqn 3, Fig 1), such that a smaller n/N_e ratio results in a higher θ value and a larger weight of the prior, both resulting in a larger penalty for sibship assignments. Like any priors in the Bayesian literature, my prior is informative and somewhat subjective, and I do not claim it is optimal. However, my simulations using many different parameter combinations (n, N_e, L, A, …) and species models verify that the prior works very well, making the SF estimator essentially unbiased and much more accurate than other single sample estimators when marker information is scarce. On the other hand, the SF estimator becomes increasingly independent of the prior with an increase in marker information (L, A) and sample size (n), as expected.

A difficulty in applying the SF estimator occurs when one has no prior information about the effective size of the population. Prior N_e becomes irrelevant when either n is large (relative to the unknown N_e, Fig 1) or marker information is ample (Fig 5). Otherwise, the N_e estimate from SF method increases with an increasing prior N_e (Fig 5). The method provides underestimates and overestimates when prior N_e is smaller and larger than the actual N_e, respectively. In the case where no prior N_e information is available, a number of different prior N_e values can be used in the SF estimator and the prior that results in the same estimate is the most likely effective size. This approach was applied to the grizzly bear data (Fig 11) and yielded sensible results. Alternative and computationally simpler approaches are to use the default sibship prior and LD estimate of N_e as prior, as shown in Fig 5.

Accuracy of single sample estimators: Across many scenarios involving population (e.g. actual N_e, genome size), sample (e.g. numbers of loci and individuals) and marker (e.g. polymorphisms and allelic dropout rates) properties, the SF and LD estimators are less biased and more accurate than HE and MC estimators by roughly one order. Relatively, HE is better than MC. Both provide very poor estimates of N_e, especially when the actual population is not very small (Fig 3). MC always underestimates N_e substantially, except for very small populations (say, actual N_e <30). The bias of HE depends critically on the ratio n/N_e, and decreases with an increasing n/N_e (Fig 4). My simulation results confirm previous studies (Pudovkin *et al.* 1996; Luikart & Cornuet 1999) that HE has some value in applications to very small populations, or in the situation where sample size is larger than the actual N_e. It can also provide almost unbiased N_e estimates when many markers (Fig 7) are used. However, even in these situations, the precision of this estimator is low, much lower than that

of the LD and SF estimators. Furthermore, its high vulnerability to non-random mating and imperfect markers (e.g. allelic dropouts, null alleles) renders it impractical.

Most previous simulation studies on N_e estimators considered small to medium sized populations with $N_e \leq 100$ (e.g. Wang 2001; Tallman *et al.* 2004; England *et al.* 2006; Waples 2006; Wang 2009; Waples & Do 2008, 2010; Waples & England 2011; Waples *et al.* 2014). This is understandable because current estimators rely on information on inbreeding and genetic drift, and such information is strong relative to sampling noises only when populations are small and sampling intensity is high (i.e. n and L are big). With the rapid developments in molecular techniques, many more markers can be genotyped for a large number of individuals at ease. As a result, we have now the capacity to study big populations, despite of their weak signals of inbreeding and genetic drift. The present simulation study is perhaps the first to explore the possibility of estimating the contemporary N_e of large populations from marker data. My simulation results are highly encouraging (Fig 3, 4), and the SF and LD estimators can provide reasonably good N_e estimates of very big populations (N_e ~30000) by nowadays typical sampling efforts (n=100, L=20) in genotyping microsatellites. With a higher sampling intensity in terms of sample sizes of individuals (n) and markers (L), the two estimators can be applied to even larger populations to obtain high-quality estimates of N_e.

My simulations showed that, over many different parameter combinations, the SF estimator is always more precise and thus more accurate than the LD estimator. The performance advantage of the SF estimator remains, albeit reduced, when the LD estimates are used as the prior N_e values (Fig 5). In the case of no prior N_e information and the LD estimate is deemed unreliable, a good alternative is to use the default prior (Fig 5). When marker information (L, A) is sufficiently high (Fig 5), or sample size is large relative to the actual N_e (Fig 1), the prior has little effect on the quality of SF estimator. A survey of 89 studies published in the journal Conservation Genetics in 2014 showed that, on average, 12 microsatellites are used (Vilas *et al*. 2015). At this level of marker information, SF could rely on the prior to some extent (Fig 5), depending on the ratio n / N_e.

The extent of linkage disequilibrium observed in a population is determined by the drift occurred over a number of previous generations (Hill 1981). The LD estimate of N_e should, therefore, reflect the average effective size of the population at the parental and a number of earlier generations (Wang 2005), and should overestimate (underestimate) parental N_e if it is greatly decreased (increased). My simulations confirmed that the parental N_e is indeed overestimated and underestimated by the LD estimator when there is a sudden

decrease and increase in N_e at the parental generation, respectively (Fig 6). However, the bias of LD estimator is not substantial in both cases. However, with linked markers, LD estimator could reflect the effective size many generations before the sampling point, and thus could become more biased (improper) as an estimate of the parental N_e when the population has a fluctuating size and/or breeding system. On the other hand, the bias of LD as an estimator of parental N_e as observed in Fig 6 would be even smaller if the population demographic change occurs earlier than the parental generation.

This study focused on the bias, precision and accuracy of different estimators, and did not compare their computational time. LD, HE and MC are (allele frequency) moment-based methods and are very simple to calculate, taking usually just a couple of seconds to finish an analysis. With an increase in the number of markers L, MC estimator becomes slower. Even when L =1000, however, an MC analysis takes just a few minutes. In contrast, SF method uses a likelihood based sibship assignment analysis to obtain sibship frequencies. The analysis is computationally intensive due to the complexity of computing the likelihood of a sibship configuration and the algorithm, simulated annealing, used to maximize the likelihood (Wang & Santure 2009). For the analysis of a typical dataset in this simulation study, the SF method takes about 20 minutes with the default parameter settings, except for the prior. The computational time increases quickly with sample size, and can take days or weeks to finish a run for a large dataset with many individuals and marker loci with genotyping errors. However, in the case of very large dataset, a likelihood score method (Wang 2012) can be used to quickly assign sibships with reasonable accuracy.

A single sample estimator based on multiple summary statistics (Tallman *et al.* 2004) is not used in my comparison study. The estimator has been implemented in a software package, OneSamp (Tallman *et al.* 2008). However, it is available only as a web-based program and thus is infeasible for a simulation study in which hundreds of thousands of datasets must be analysed. Conceptually, this estimator does not estimate the contemporary N_e because it uses some summary statistics (e.g. number of alleles per locus) that are determined by N_e over an evolutionary time scale (Wang *et al*. 2016). Therefore, this estimator may not be comparable with other single-sample estimators in practice, because no real populations have a stable size in such a long time of many generations.

Robustness: Each N_e estimator is based on a genetic model under a number of simplifying assumptions (briefed in Table 2), such as discrete generations, random mating, an isolated population with no immigration, no linkage among markers, and no genotyping error of markers. Violation of these assumptions may have a highly variable effect on different

estimators. Waples and coworkers (Waples & England 2011; Waples *et al.* 2014) investigated the robustness of LD estimator applied to populations with overlapping generations and to populations receiving immigrants. In their simulations with immigration, samples containing first-generation immigrants (i.e. both parents of an immigrant are from the source population) were used in LD estimator. They found that LD estimator is surprisingly robust to immigration, providing good estimates of local effective population size except when migration rate is greater than 5~10% (Waples & England 2011). My simulations (Wang 2009, and results not shown in this study) showed that the SF estimator is unaffected by immigration when it is applied to a sample of individuals containing no first-generation immigrants, irrespective of the migration rate. Immigrants in the parental and more remote generations do not affect the SF estimator, because they do not affect the rate of inbreeding or genetic drift of the local population. First-generation migrants do affect the sibship frequencies, but they can be avoided by careful experimental and sampling designs (e.g. sampling before the life stage in which migration occurs) or by the detection and elimination using the multilocus genotype data (Pritchard *et al.* 2000; Rannala & Mountain 1997).

When a single cohort is sampled from a population with overlapping generations and is analysed by the four estimators, a statistic called effective number of breeders (Pudovkin 1996; Nomura 2008; Waples & Antao 2014; Waples *et al.* 2014), N_b, rather than effective size, N_e, is obtained. Conceptually, N_b is different from the effective population size per generation (N_e) or per year (N_y, or annual effective population size, Hill 1979). It summarises partially the effects of the sizes of age classes and the individual variation in reproductive contributions within and among age and sex classes on the stochastic processes of inbreeding and genetic drift in a population with overlapping generations. However, the effects of life span and other factors, which are also relevant in determining the stochastic process, are not reflected in N_b. The LD estimator is now almost routinely used to estimate N_b for populations with overlapping generations (e.g. Duong *et al.* 2013; Whiteley *et al.* 2014). Simulations have also been conducted to investigate what the LD estimates really are when calculated from a single cohort sample and mixed cohort samples (e.g. Robinson & Moyer 2013; Waples *et al.* 2014) from a population with overlapping generations. For understanding the genetic stochasticity of a population with overlapping generations, estimators of N_e and generation interval dedicated for such populations, such as that based on parentage analysis (Wang *et al.* 2010) should be applied. A recent study applying the estimator to a grizzly bear population monitored over more than 50 years (Kamath *et al*. 2016) showed that both generation interval and N_e have been increasing in this time period.

An advantage of the SF estimator is that it can be applied to non-random mating populations. Unlike other single sample estimators which invariably assume random mating, SF allows for non-random mating by using the parameter α, equivalent to Wright's F_{IS}, in the estimation equations (e.g. eqn (4)). This parameter measures the departure of genotype frequencies from Hardy-Weinberg equilibrium caused by non-random mating, and can be estimated from the same multilocus genotype data as sibship frequencies (Wang 2009). In general, however, the effect of non-random mating on N_e is small in dioecious outbreeding species. For monoecious species, however, N_e can be much reduced by selfing when it occurs at a substantial rate. My simulations showed that the SF estimator, (4), can be applied to obtain accurate N_e estimates when selfing rate is not very high. Otherwise, sibship will be over-assigned and N_e underestimated. In such cases, a correction factor can be used to remove the bias (Fig. 8).

With the rapid development of sequencing technology, genome-wide SNPs are increasingly used in molecular ecology and conservation genetics studies. These markers are necessarily linked when numerous. However, their linkage relationships (genetic map distances) might be poorly known. This study is the first to investigate the performances of LD and other single sample estimators when they are applied to many linked SNPs under the assumption of no linkage. As is expected, SF, HE and MC are little affected by linkage, but LD estimator underestimates N_e severely when the linkage among SNPs is strong (Fig 7). More work is needed to evaluate the performance of Hill's (1981) original LD estimator (which allows for linkage) when the genetic map distances among SNPs are known and are used in the estimator.

The SF estimator is also robust to imperfect data. My simulations confirm (Fig 10) that SF estimator is robust to allelic dropouts, and can yield fairly good estimates of N_e even when dropouts occur at a much higher rate than that assumed in the analysis. In contrast, LD and HE estimators are vulnerable to dropouts. Although MC is little affected by dropouts, it almost always underestimates N_e substantially and is much less accurate than other estimators, no matter allelic dropouts are present or not. I also simulated null alleles (data not shown), and obtained results and reached conclusions similar to those for allelic dropouts regarding the relative performances of different estimators. Sved *et al*. (2013) showed that LD estimator is sensitive to the presence of null alleles, which leads to an underestimation of N_e. They showed, however, that a permutation analysis could be used to correct for the inflated linkage disequilibrium due to null alleles, and thus to obtain essentially unbiased estimates of N_e. They did not however evaluate the accuracy of LD estimator as measured by

RMSE in the presence of null alleles with and without applying the permutation correction. I surmise the same permutation procedure could also be used to correct for the bias caused by allelic dropouts. Further work is needed in this area.

Sibship analysis (and thus SF estimator) requires prior information about the mating system. When a species is designated as polygamous (monogamous) for females, then maternal half sibship is assumed to be present (absent) among sampled individuals and such relationship will (not) be assigned in a sibship analysis. This is also true for male mating system and paternal half sibship. Therefore, mis-specifying polygamy as monogamy when there exist a lot of half siblings in the sample will lead to no half sibship assignments, an underestimation of sibship frequency, and thus an overestimation of N_e, no matter how many markers are used in the sibship analysis. In contrast, mis-specifying monogamy as polygamy when no half siblings exit in the sample only leads to a possible reduction in the power (accuracy) of a sibship analysis. With sufficient marker information (say, 20 microsatellites), however, full sibship should be accurately inferred and no half sibship should be assigned when monogamy is mis-specified as polygamy (Wang 2004). In the absence of any information, the mating system should be better designated as polygamous for both males and females. Note the LD method also needs mating system information to calculate the expected linkage disequilibrium correctly (Hill 1981).

Application scopes: Compared with LD and other single sample estimators, the SF estimator has a much broader application scope and could handle different species models and markers. For example, SF estimator applies to dioecious diploid species with random or non-random mating, to haplodiploid species (Fig 9), and to monoecious species with mixed selfing and outbreeding (Fig 8). It can also be easily adapted to apply to polyploid species, because the current likelihood sibship assignment method can yield accurate sibship assignments for polyploid species (Wang & Scribner 2014). In terms of markers, SF method applies to microsatellites, many linked SNPs without knowing linkage relationships, and dominant markers. Considering the wider application scope, more robustness (e.g. to migration, non-random mating and genotyping errors) and higher accuracy of the SF estimator, I recommend its wide use in practice.

Acknowledgements

I thank Robin Waples, two anonymous referees and the editor, Nick Barton, for their valuable comments on an earlier version of this manuscript.

Literature

Anderson EC (2005) An efficient Monte Carlo method for estimating N_e from temporally spaced samples using a coalescent based likelihood. *Genetics*, **170**, 955-967.

Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C, Taberlet P (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, **13**, 3261-3273.

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

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

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

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 (N_e) from genetic data. *Molecular Ecology Resources*, **14**, 209-214.

Duong TY, Scribner KT, Forsythe PS, Crossman JA, Baker EA (2013) Interannual variation in effective number of breeders and estimation of effective population size in long-lived iteroparous lake sturgeon (*Acipenser fulvescens*). *Molecular Ecology*, **22**, 1282-1294.

England PR, Cornuet JM, Berthier P, Tallmon DA, Luikart G (2006) Estimating effective population size from linkage disequilibrium: severe bias in small samples. *Conservation Genetics*, **7**, 303-308.

Ewens WJ (1972) The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, **3**, 87-112.

Fisher RA (1930) *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

Haldane JBS (1919) The combination of linkage values, and the calculation of distances between the loci of linked factors. *Journal of Genetics*, **8**, 99-309.

Haldane JBS (1924) A mathematical theory of natural and artificial selection. II. The influence of partial self-fertilisation, inbreeding, assortative mating, and selective fertilisation on the composition of Mendelian populations, and on natural selection. *Biological Reviews*, **1**, 158-163.

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. *Genetical Research*, **38**, 209-216.

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

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. *Molecular Ecology*, 24, 5501-5521.

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. *Conservation Genetics*, **11**, 355-373.

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

Nomura T (2008) Estimation of effective number of breeders from molecular coancestry of single cohort sample. *Evolutionary Applications*, **1**, 462-474.

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.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.

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.

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 9197-9201.

Robertson A (1965) The interpretation of genotypic ratios in domestic animal populations. *Animal Production*, **7**, 319-324.

Robinson JD, Moyer GR (2013) Linkage disequilibrium and effective population size when generations overlap. *Evolutionary Applications*, **6**, 290-302.

Schwartz MK, Tallmon DA, Luikart GH (1999) DNA-based N_e estimation: many markers, much potential, uncertain utility. *Animal Conservation*, **2**, 320-322.

Schwartz MK, Luikart G, Waples RS (2007) Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution*, **22**, 25-33.

Sved JA, Cameron EC, Gilchrist AS (2013) Estimating effective population size from linkage disequilibrium between unlinked loci: theory and application to fruit fly outbreak populations. *PloS one*, **23**, e69078.

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) ONeSamp: a program to estimate effective population size using approximate Bayesian computation. *Molecular Ecology Resources*, **8**, 299-301.

Vilas A, Pérez-Figueroa A, Quesada H, Caballero A (2015) Allelic diversity for neutral markers retains a higher adaptive potential for quantitative traits than expected heterozygosity. *Molecular Ecology*, **24**, 4419-4432.

Wang J (2001) A pseudo-likelihood method for estimating effective population size from temporally spaced samples. *Genetical Research*, **78**, 243-257.

Wang J (2005) Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society, Series B: Biological Sciences*, **360**, 1395-1409.

Wang J (2009) A new method for estimating effective population size from a single sample of multilocus genotypes. *Molecular Ecology*, **18**, 2148-2164.

Wang J (2012) Computationally efficient sibship and parentage assignment from multilocus marker data. *Genetics*, **191**, 183-194.

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

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

Wang J, Scribner K (2014) Parentage and sibship inference from markers in polyploids. *Molecular Ecology Resources*, **14**, 541-553

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, Santiago E, Caballero A (2016) Prediction and estimation of effective population size. Heredity (Submitted).

Wang J, El-Kassaby YA, Ritland K (2012) Estimating selfing rates from reconstructed pedigrees using multilocus genotype data. *Molecular Ecology*, **21**, 100-116.

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.

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

Waples RS (2006) A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conservation Genetics*, **7**, 167-184.

Waples RS, Do C (2008) LdNe: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources*, **8**, 753-756.

Waples RS, Do C (2010) Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary Applications*, **3**, 244-262.

Waples RS, Antao T (2014) Intermittent breeding and constraints on litter size: consequences for effective population size per generation (Ne) and per reproductive cycle (Nb). *Evolution*, **68**, 1722-1734.

Waples RS, Antao T, Luikart G (2014) Effects of overlapping generations on linkage disequilibrium estimates of effective population size. *Genetics*, **197**, 769-780.

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, Waples RK (2011) Inbreeding effective population size and parentage analysis without parents. *Molecular Ecology Resources*, **11**, 162-171.

Whiteley AR, McGarigal K, Schwartz MK (2014) Pronounced differences in genetic structure despite overall ecological similarity for two Ambystoma salamanders in the same landscape. *Conservation Genetics*, **15**, 573-591.

Whitlock M, Barton N (1997) The effective size of a subdivided population. *Genetics*, **146**, 427-441.

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

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

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

J. Wang is interested in developing population genetics models and methods of analysis of empirical data to address issues in evolutionary and conservation biology.

Data accessibility

The computer program for simulating genotype data under different parameter combinations, and for estimating N_e from different single sample estimators: Dryad DOI: ###.

The empirical dataset was retrieved from Dryad with DOI: http://dx.doi.org/10.5061/dryad.s6764.

Table 1 Parameter combinations in simulations

Simulation	Focal parameter (values)	Fixed parameter values						Result Fig.
		Species	N_1, N_2 (N)	N_e	L	A	n	
1	L(5, 10, 15, 20, 25, 30, 35, 40)	DD	50, 50	100	-	10	50	2
2	n(20, 40, 80, 160, 320)	DD	250, 250	500	20	10	-	3
3	N_e (10, 50, 250, 1250, 6250, 31250)	DD	N_e/2, N_e/2	-	20	10	100	4
4	Prior N_e (5, 10, 20, 40, 80, 160, 320, LD estimate)	DD	25, 25	50	10, 50	10	50	5
5	G (-40, -20, 10, 0, 10, 20, 40)	DD	50, 50	100× (1+G)	20	10	100	6
6	M (1, 2, 4, 8, 16, 32, 64)	DD	25, 25	50	1000	2	50	7
7	s (0, 0.1, 0.2, 0.4, 0.8, 0.98)	MD	(100)	100, 95, 90, 80, 60	10	10	50	8
8	L (5, 10, 20, 40)	DH	50, 50	75	-	10	50	9
9	D (0, 0.025, 0.05, 0.1, 0.2, 0.4)	DD	25, 25	50	10	10	50	10

The three species models are dioecious diploid (DD), dioecious haplodiploid (DH), and monoecious diploid (MD). The column headed by "N_1, N_2 (N)" gives the numbers of males and females in DD models and the number of individuals in MD models. N_e is the theoretical effective size. Except when explicitly explored, the simulated population is assumed isolated, constant in size (no bottleneck), and the markers are assumed unlinked (i.e. genetic map length of the genome M=∝). G is the percentage change in parental population size (so that N_i(1+G) is the number of breeders of sex i, while in other generations the number is N_i). D is the rate of allelic dropouts at each locus. L is the number of markers. A is the number of alleles per locus. n is sample size (number of individuals). s is selfing rate.

Table 2 Comparison of single-sample estimators of N_e

Method	Information used	Key assumptions	Strengths	Weaknesses	Software and reference
HE	Heterozygosity excess	Random sampling; An isolated random mating population; Diploid; Codominant markers; No allelic dropouts; No null alleles	Simple computation; Nearly unbiased with $n/N_e>1$; Robust to linkage	Imprecise; Highly biased with $n/N_e<1$, non-random mating, allelic dropouts, or null alleles; Unsuitable for dominant markers, and for haplodiploid species	NeEstimator, Do *et al*. 2014; Colony, Jones & Wang 2010; Nb_HetEx, Zhdanova & Pudovkin 2008
MC	Molecular coancestry	Random sampling; An isolated random mating population; Diploid; Codominant markers	Simple computation; Robust to allelic dropouts, null alleles, and linkage	Highly biased and inaccurate	NeEstimator, Do *et al*. 2014
LD	Linkage disequilibrium	Random sampling; An isolated random mating population; Diploid; Codominant markers; No allelic dropouts; No null alleles; No linkage	Simple computation; Accurate when assumptions are met	Inaccurate with linkage, non-random mating, population structure, allelic dropouts, null alleles. Unsuitable for haplodiploid species; Limited ability for bottleneck detection	NeEstimator, Do *et al*. 2014; LDNE, Waples & Do 2008
SF	Sibship frequency	Random sampling; Diploid or haplodiploid species	Accurate; Wide application scope (non-random mating; subdivided population; diploid and haplodiploid species; dominant and codominant markers); Highly robust to allelic dropouts and null alleles, and to linkage	Highly computational demanding; Sensitive to improper priors when marker information is scarce and n/N_e is small	Colony, Jones & Wang 2010

Note, the LD method refers to the improved estimator by Waples (2006) and Waples & Do (2008), which assumes unlinked codominant markers. *n*, number of individuals in a sample.

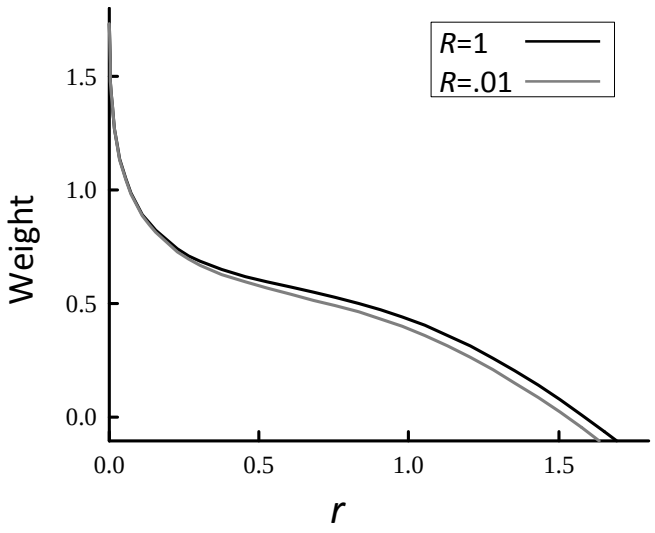

Fig. 1 Weight of the sibship assignment prior as a function of r (sample size n to prior N_e ratio) and prior sex ratio R.

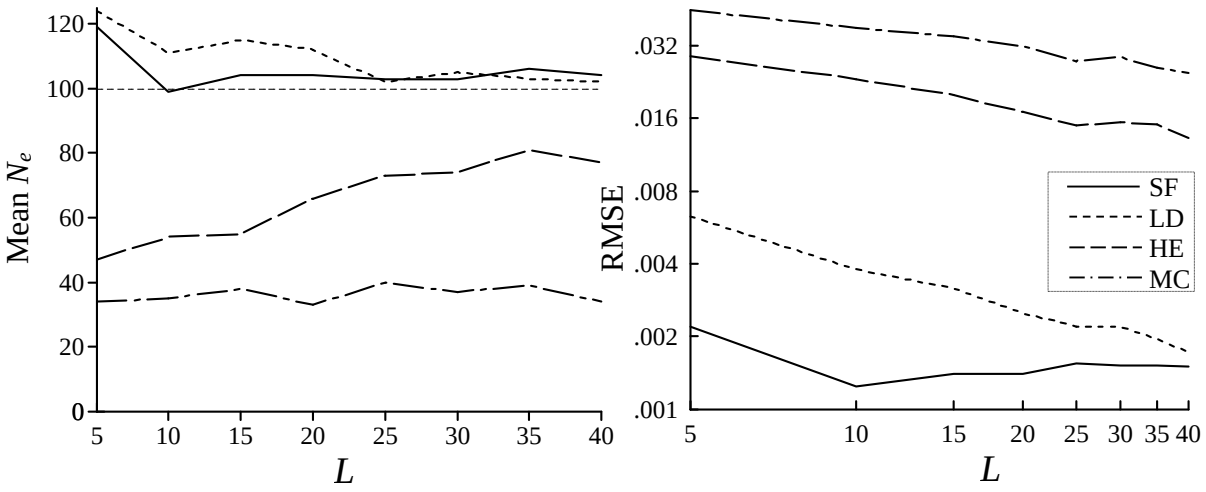

Fig. 2 Mean and RMSE of N_e estimates as a function of the number of loci, L. An isolated population was simulated under a DD model for a variable number of L unlinked loci, and the other parameters being fixed at A=10, N_1=N_2=50, n=50. The simulated N_e is 100. Note both x and y axes are in log scale in the plot of RMSE.

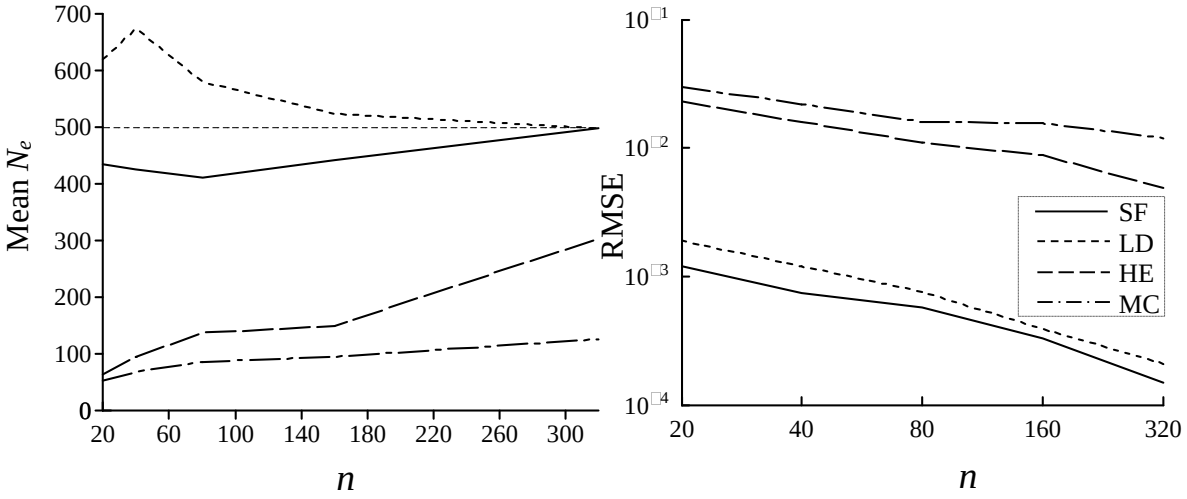

Fig. 3 Mean and RMSE of N_e estimates as a function of sample size, n. An isolated population was simulated under a DD model for a variable sample size of n individuals, and the other parameters being fixed at A=10, N_1=N_2=250, L=20. The simulated N_e is 500. Note both x and y axes are in log scale in the plot of RMSE.

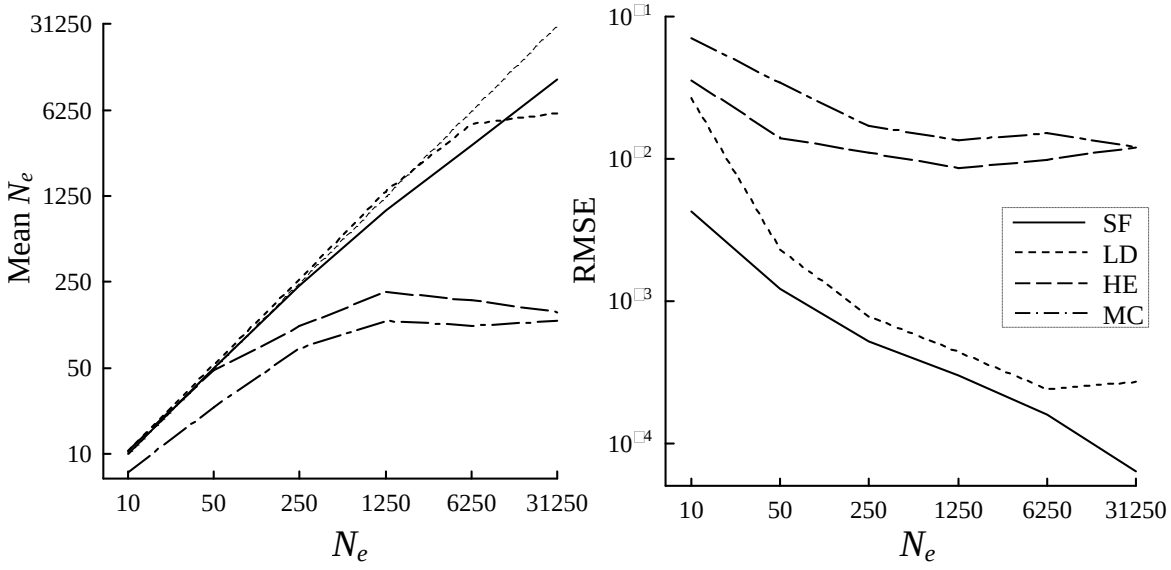

Fig. 4 Mean and RMSE of N_e estimates as a function of the actual effective population size, N_e. An isolated population was simulated under a DD model for a variable N_e (N_1=N_2= N_e / 2), and the other parameters being fixed at A=10, n=100, L=20. Note both x and y axes are in log scale in both plots of mean N_e estimates and RMSE.

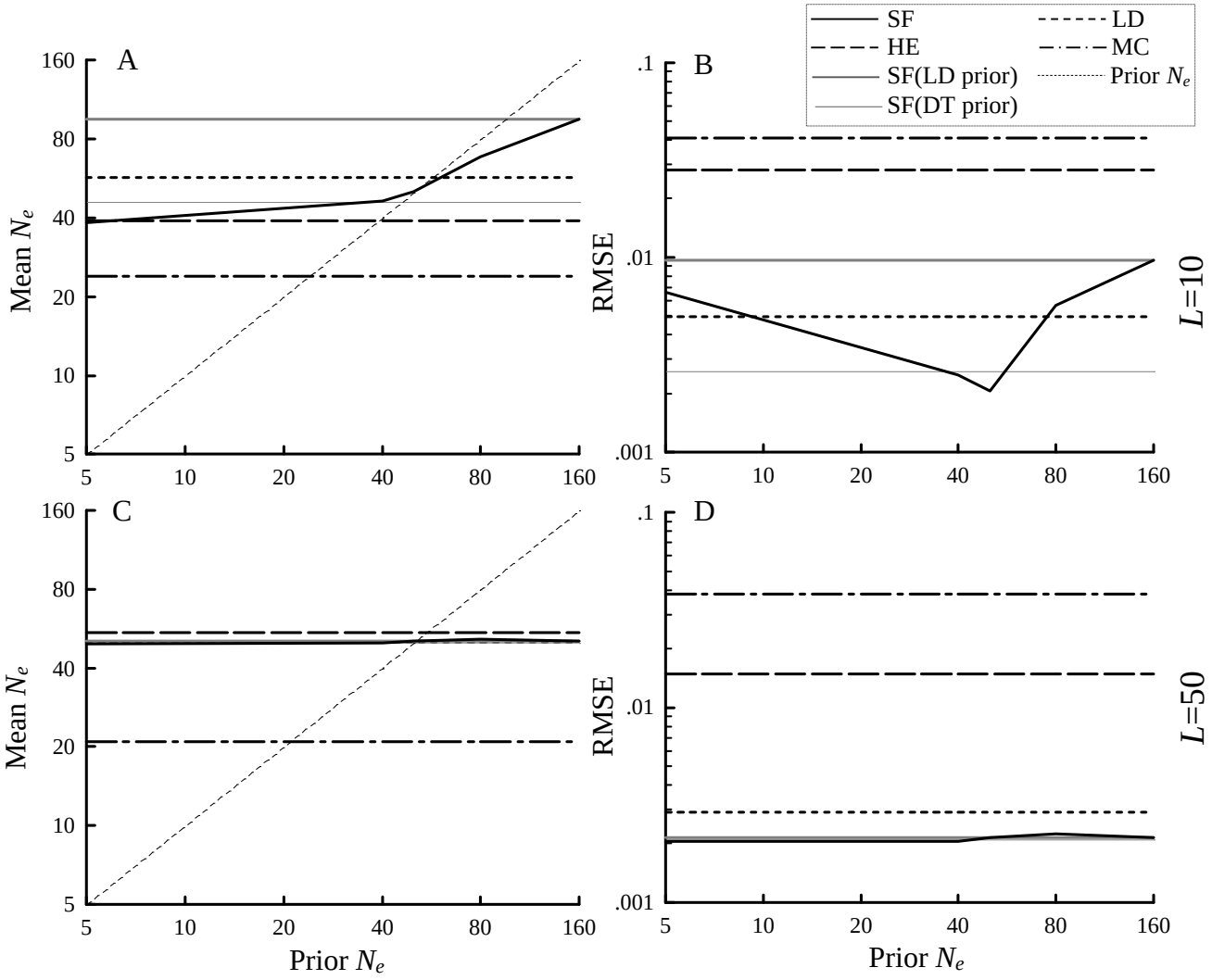

Fig. 5 Mean and RMSE of N_e estimates as a function of the prior N_e used in SF estimators. No priors were used in HE, MC and LD estimators and their estimates were invariable with prior N_e. The SF estimator used either a variable N_e value on *x* axes as priors (continuous black lines, labelled as SF), the LD estimate as priors (thick continuous grey lines, labelled as SF(LD prior)), or the default (unknown N_e) priors (thin continuous grey lines, labelled as SF(DT prior)). An isolated population was simulated under a DD model for a variable number of L= 10 or L=50 loci, and the other parameters being fixed at N_e =50 (N_1=N_2= 25), A=10, n=50. Note both *x* and *y* axes are in log scale in both plots of mean N_e estimates and RMSE, and the lines for SF, SF(LD prior) and SF(DT prior) are indistinguishable when L=50.

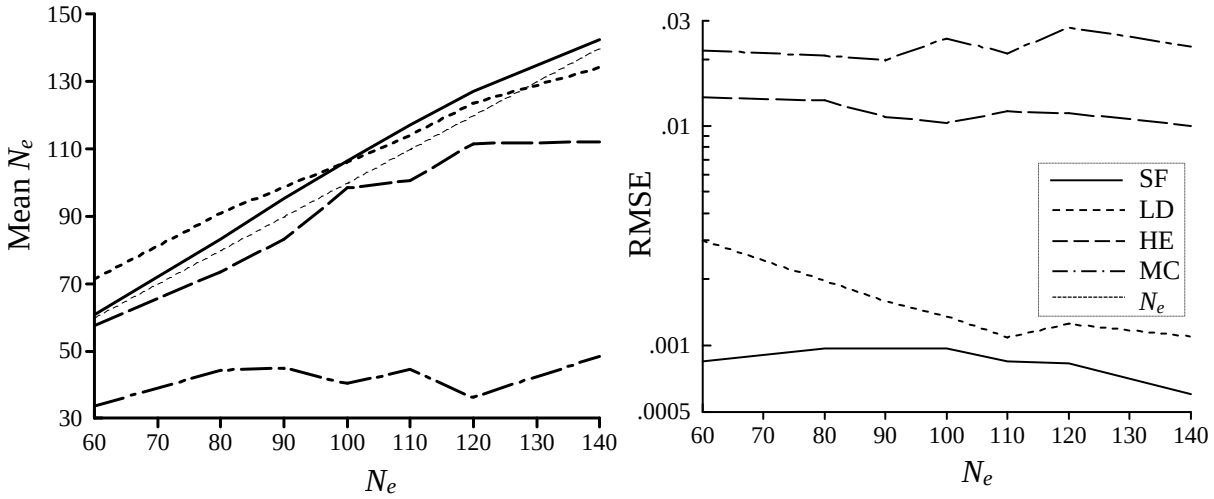

Fig. 6 Mean and RMSE of N_e estimates as a function of the actual (simulated) parental N_e. An isolated population was simulated under a DD model for a constant population size (N_1=N_2= N_e / 2=50) except in the parental generation in which the population size changes to those shown on the *x* axes. The other parameters were fixed at L=20, A=10, n=100. Note the *y* axis is in log scale in the plot of RMSE. The theoretical (simulated) parental N_e is shown in the thin dotted line in the plot of mean N_e.

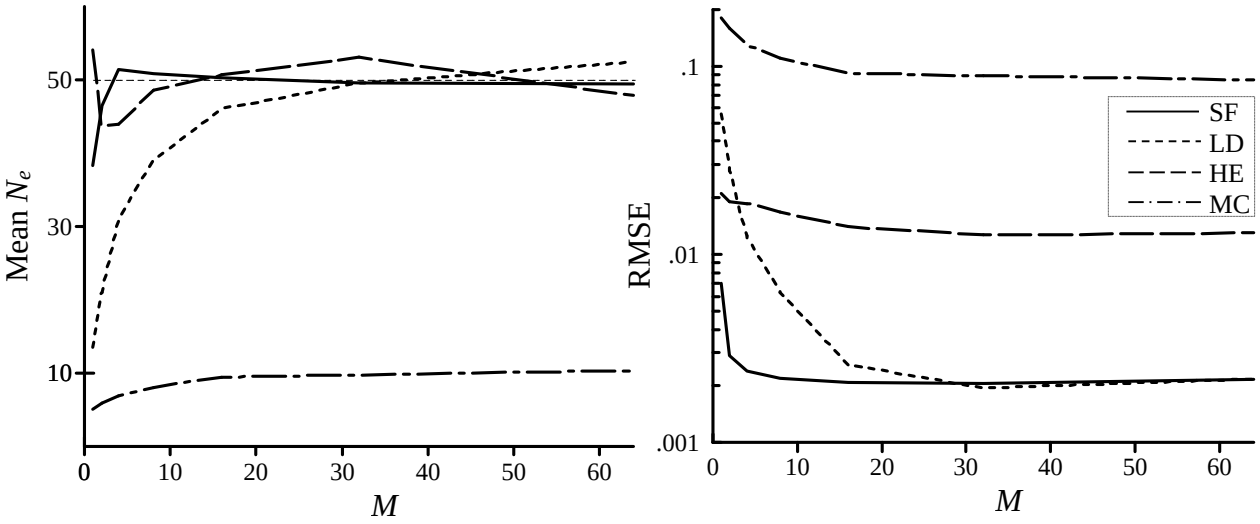

Fig. 7 Mean and RMSE of N_e estimates as a function of the map length (M, in Morgan) of the genome. An isolated population was simulated under a DD model for a variable M (x axes), and the other parameters being fixed at L=1000, A=2, N_1=N_2= N_e / 2=25, n=50. Note the y axis is in log scale in the plot of RMSE, and the simulated value of N_e =50 is shown in thin dotted line in the plot of mean N_e.

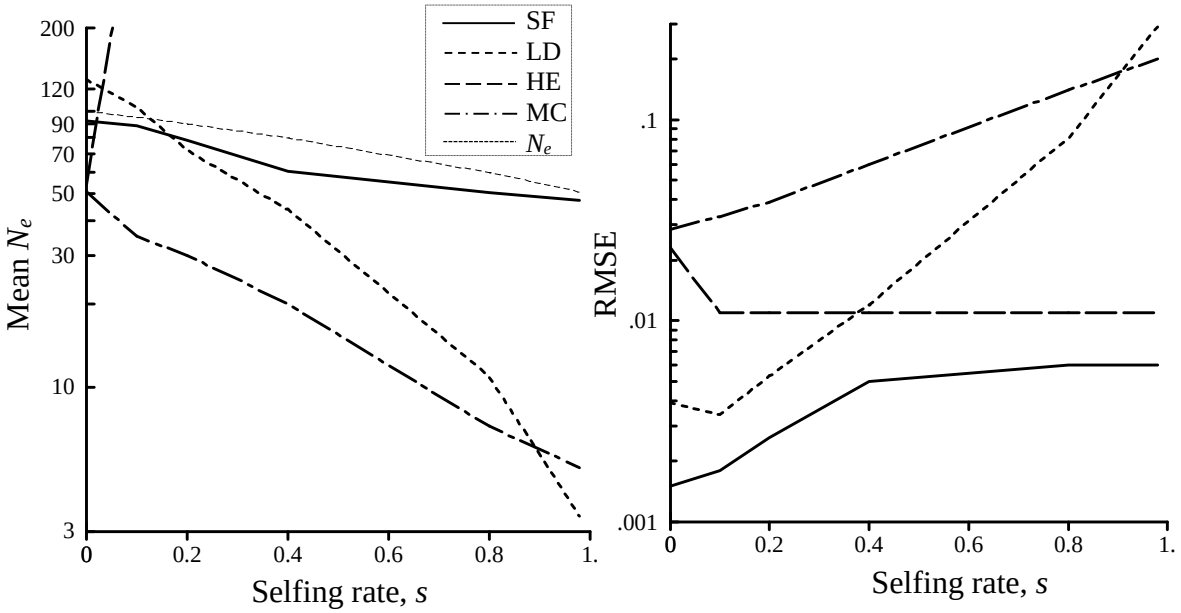

Fig. 8 Mean and RMSE of N_e estimates as a function of the actual selfing rate (s). An isolated population was simulated under a MD model for a variable rate of selfing (x axes), and the other parameters being fixed at L=10, A=10, N=100, n=50. The theoretical (simulated) N_e is shown in the thin dotted line. Note the y axes are in log scale in both plots of mean N_e estimates and RMSE.

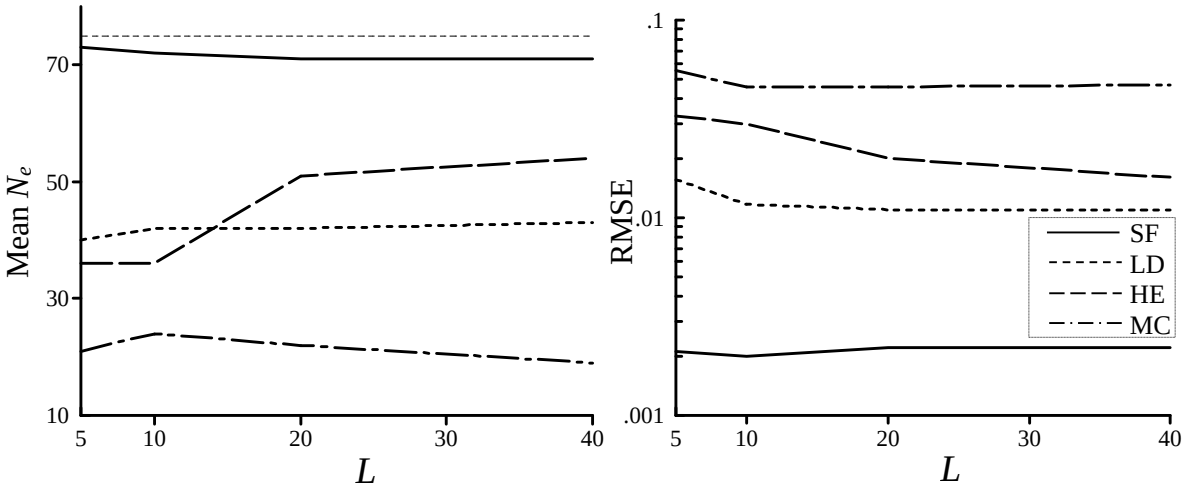

Fig. 9 Mean and RMSE of N_e estimates as a function of the number of loci (L). An isolated population was simulated under a dioecious haplodiploid (DH) model, and the other parameters being fixed at A=10, N_1=N_2=50, N_e=75, n=50. The theoretical (simulated) N_e =75 is shown in the thin dotted line. Note the y axis is in log scale in the plot of RMSE.

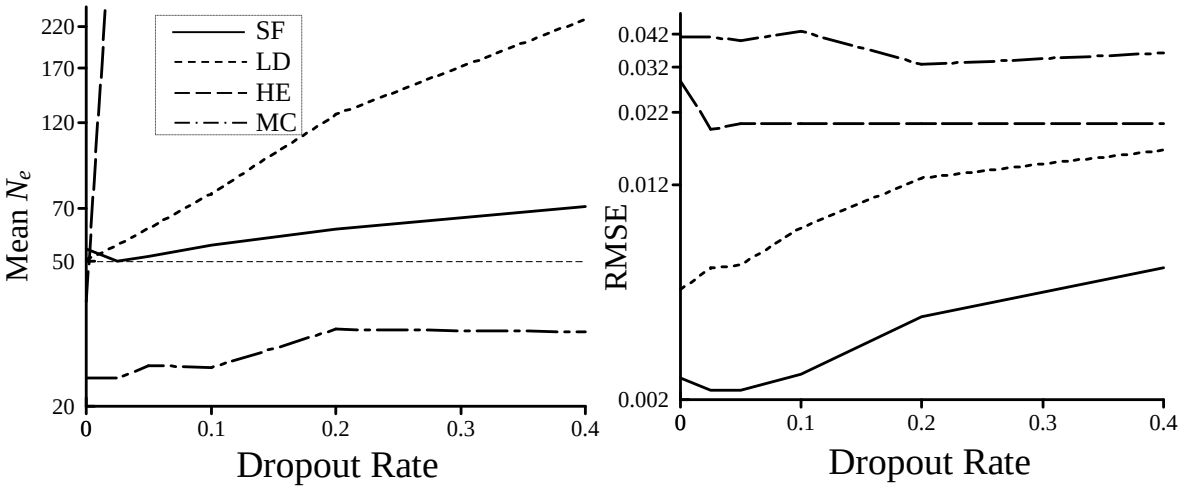

Fig. 10 Mean and RMSE of N_e estimates as a function of the rate of allelic dropouts at each locus. An isolated population was simulated under a dioecious diploid (DD) model, and the other parameters were fixed at L=10, A=10, N_1=N_2=25, N_e=50, n=50. The theoretical (simulated) N_e =50 is shown in the thin dotted line. Note the y axis is in log scale in the plot of RMSE.

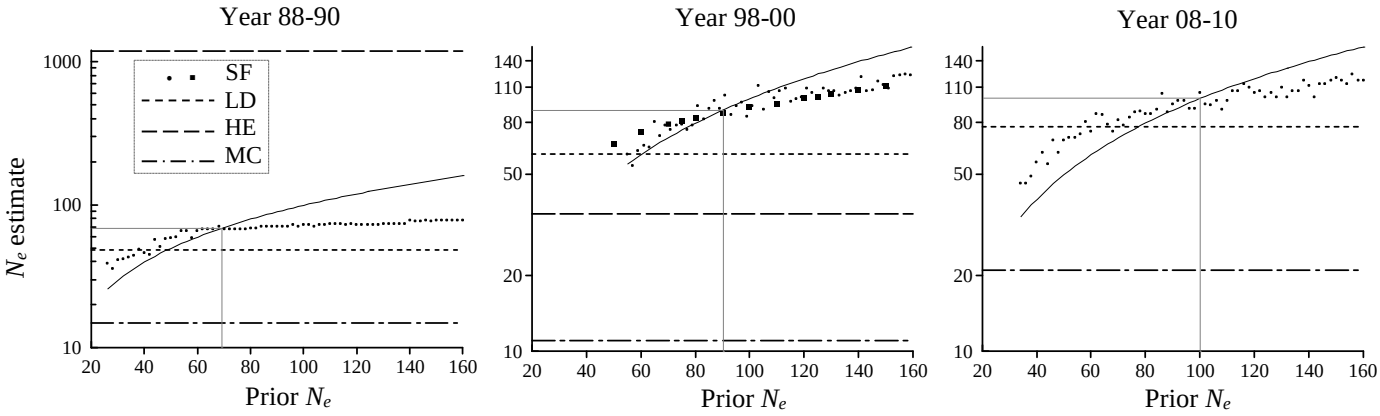

Fig. 11 Estimates of N_e as a function of prior N_e values (on the *x* axes) for the grizzle bear data in the year periods 1988-1990, 1998-1990, and 2008-2010. The sample sizes for the 3 periods are 46, 92, and 59, respectively. For each period, N_e was estimated by LD, HE, MC and SF methods. For SF, estimates (in scattered dots) were obtained by assuming different prior N_e values (plotted in the continuous thin lines). For the period 1998-1990, harmonic mean estimates obtained by SF over 100 bootstrapping samples of size 50 for each assumed prior N_e were also plotted (squares). Note the *y* axes are in log scale, and the HE estimate of N_e is infinitely large for the period 2008-2010 so it is not plotted.