Estimating the Effective Population Size from Temporal Allele Frequency Changes in Experimental Evolution

Ágnes Jónás,*,† Thomas Taus,*,† Carolin Kosiol,† Christian Schlötterer,† and Andreas Futschik†,‡,1

*Vienna Graduate School of Population Genetics, 1210 Vienna, Austria, †Institut für Populationsgenetik, Vetmeduni Vienna, 1210 Vienna, Austria, and ‡Department of Applied Statistics, Johannes Kepler Universität Linz, 4040 Linz, Austria

ORCID ID: 0000-0002-7980-0304 (A.F.)

ABSTRACT The effective population size (N_e) is a major factor determining allele frequency changes in natural and experimental populations. Temporal methods provide a powerful and simple approach to estimate short-term N_e. They use allele frequency shifts between temporal samples to calculate the standardized variance, which is directly related to N_e. Here we focus on experimental evolution studies that often rely on repeated sequencing of samples in pools (Pool-seq). Pool-seq is cost-effective and often outperforms individual-based sequencing in estimating allele frequencies, but it is associated with atypical sampling properties: Additional to sampling individuals, sequencing DNA in pools leads to a second round of sampling, which increases the variance of allele frequency estimates. We propose a new estimator of N_e, which relies on allele frequency changes in temporal data and corrects for the variance in both sampling steps. In simulations, we obtain accurate N_e estimates, as long as the drift variance is not too small compared to the sampling and sequencing variance. In addition to genome-wide N_e estimates, we extend our method using a recursive partitioning approach to estimate N_e locally along the chromosome. Since the type I error is controlled, our method permits the identification of genomic regions that differ significantly in their N_e estimates. We present an application to Pool-seq data from experimental evolution with *Drosophila* and provide recommendations for whole-genome data. The estimator is computationally efficient and available as an R package at https://github.com/ThomasTaus/Nest.

KEYWORDS effective population size; genetic drift; Pool-seq; experimental evolution

DURING experimental evolution studies, populations are maintained under specific laboratory conditions (Kawecki *et al.* 2012; Long *et al.* 2015; Schlötterer *et al.* 2015). In sexually reproducing organisms, the census population size is typically kept fixed at fairly low numbers, rarely exceeding 2000 individuals. With such small population sizes, genetic drift causes stochastic fluctuations in allele frequencies. Under neutrality, the level of random frequency changes is determined by the effec-tive population size (N_e) (Wright 1931). Furthermore, the efficacy of selection is influenced by N_e. For weakly selected alleles, the probability of fixation is directly proportional to the product of N_e and the intensity of selection (Fisher 1930; Kimura 1964). As changes in allele frequency are greatly affected by the population size, it is fundamental to estimate N_e accurately to understand molecular variation in experimental evolution studies.

Krimbas and Tsakas (1971) estimated N_e using the standardized variance of allele frequency (*F*, see also Falconer and Mackay 1996) from longitudinal samples in natural populations of olive flies. As *F* was calculated from these samples, they accounted for the sampling variance that also contributed to the true allele frequency variance. This approach was further improved and used by several authors (Nei and Tajima 1981; Pollak 1983; Waples 1989; Jorde and Ryman 2007).

With the widespread availability of powerful computers, also maximum-likelihood-based methods became popular (Williamson and Slatkin 1999; Anderson *et al.* 2000; Wang 2001; Hui and Burt 2015) in addition to the moment-based

approaches discussed above. Although these methods show less bias than the moment-based approaches (Wang 2001), they are still computationally demanding, in particular for the large numbers of markers typically obtained with novel sequencing technologies (Foll *et al.* 2015).

Estimating N_e with temporal methods requires samples collected at least at two time points. Alternative methods that use only a single time point are based on linkage disequilibrium (LD) (Hill 1981; Waples and Do 2008, 2010; Waples and England 2011), heterozygote excess (Pudovkin *et al.* 1996), molecular coancestry (Nomura 2008), sibship frequencies (Wang 2009, 2013), or combinations of summary statistics using approximate Bayesian computation (Tallmon *et al.* 2008). LD-based methods are widespread but require haplotype or unphased diploid genotype information, which limits their applicability.

Although the cost for sequencing has dropped considerably, the separate sequencing of thousands of individuals in replicate populations in experimental evolution studies is still out of reach. Sequencing samples in pools (Pool-seq) can provide a cost-effective alternative (Schlötterer *et al.* 2014). Pool-seq has also been shown to outperform individual-based sequencing in estimating allele frequencies and inferring population genetic parameters under several conditions (Futschik and Schlötterer 2010; Zhu *et al.* 2012; Gautier *et al.* 2013). For these reasons, Pool-seq has become the basis of many experimental evolution "evolve and resequence" (E&R) studies (Turner *et al.* 2011; Schlötterer *et al.* 2015). Following the emergence of E&R, many population genetic estimators have been adjusted to handle the properties of Pool-seq data (Futschik and Schlötterer 2010; Kofler *et al.* 2011a,b; Kolaczkowski *et al.* 2011; Boitard *et al.* 2013; Ferretti *et al.* 2013). To the best of our knowledge, no N_e estimators have been developed so far that properly deal with Pool-seq data.

In this article, we present a novel temporal method to estimate N_e from pooled samples. We show that previously proposed estimators can lead to substantial bias, as they neglect the variance component due to pooled sequencing. We introduce a new model accounting for the two-step sampling process associated with Pool-seq data. In the first sampling step individuals are drawn from the population to create pooled DNA samples. In the second step, pooled sequencing is modeled as binomial sampling of reads out of the DNA pool. We show on simulated data that our method outperforms standard temporal N_e estimators. For real data, we also suggest to use a segmentation algorithm, to partition the genome-wide sequence data into stretches of DNA with significantly different N_e estimates. Finally, we present an application to a genome-wide experimental evolution data set from *Drosophila melanogaster* (Franssen *et al.* 2015).

Materials and Methods

Sampling schemes

Nei and Tajima (1981) investigated the sampling properties of temporal N_e estimators and proposed two different sampling schemes. Under the first scheme (plan I), individuals are either sampled after reproduction or returned to the population after their genotypes have been examined. In contrast, under the second scheme (plan II) sampling takes place before reproduction and the sampled individuals are permanently removed from the population and their genotypes do not contribute to the next generation. By assuming different sampling distributions, they derived separate N_e estimators under sampling plans I and II.

Waples (1989) unified the calculations under the two plans by assuming binomial sampling out of an infinitely large parental gamete pool for both sampling schemes. He concluded that the measure of variance under the two sampling plans differs only in a covariance term. For plan I, there is a positive correlation between allele frequencies sampled t generations apart because they are both derived from the same population at generation 0. In contrast, for plan II, the initial sample and individuals contributing to the next generation can be considered as independent binomial samples; thus sample frequencies at generations 0 and t are uncorrelated.

For a typical E&R study, outbred experimental populations are created by mixing a large number of inbred lines (*e.g.*, Turner and Miller 2012; Huang *et al.* 2014; Franssen *et al.* 2015). The populations are then propagated under the desired experimental conditions while keeping the census size of the population controlled through time (Figure 1). However, the experimenter has no direct influence on the effective population size, which is in general lower than the census size. In E&R studies with *Drosophila*, the census size rarely exceeds some hundreds of individuals, and sampling usually takes place after reproduction according to plan I. For organisms maintained at larger sizes, such as yeast, the sample for genetic analysis is not returned to the population (Burke *et al.* 2014). Plan II applies to such cases.

In E&R studies, sampled individuals are often pooled together for DNA extraction (Schlötterer *et al.* 2014). The size of the pool can be as large as the whole population. Depending on the experimental design, it is also possible that only a fraction of the population is sequenced, for instance, only females (Tobler *et al.* 2014; Franssen *et al.* 2015). Pooled individuals are used to create DNA libraries, which are, in turn, subjected to high-throughput sequencing.

We consider two separate sampling steps when estimating N_e from Pool-seq samples (Figure 1). In the first step, we model the sampling of individuals out of the population. This can take place according to either plan I or plan II. In the second step, we model the sequencing of a DNA pool by drawing reads at random with replacement from the first-step sample. The allele frequency variance inferred from the sample is corrected for the additional variance coming from the two-step sampling and used for estimating N_e.

Notation

We assume that the experimental population is propagated at a constant census size N and that $N \geq N_e$. We use genome-wide single-nucleotide polymorphism (SNP) data sampled
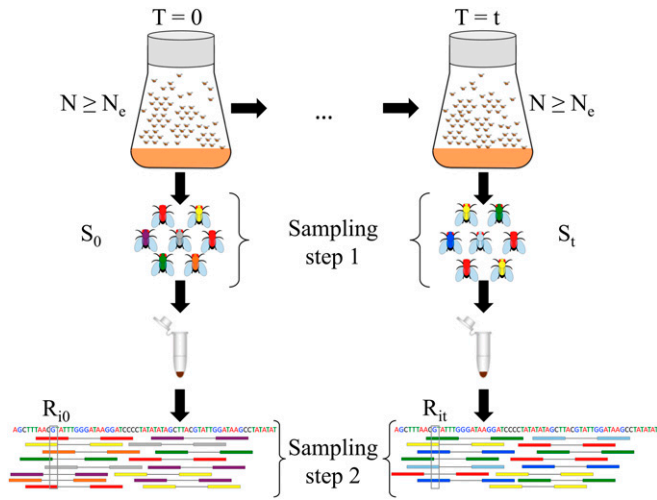
Figure 1 Two-step sampling in experimental evolution with *Drosophila*. In E&R studies, populations are propagated at a census size N defined by the experimenter, which is in general larger than the effective population size N_e. Using temporal methods, N_e can be estimated from the variance in allele frequency between samples taken t generations apart. To get an accurate representation of allele frequencies in population genetic studies, a large number of individuals S_j ($j \in \{0, t\}$) are sampled and pooled. Sampling can take place according to sampling plan I or II based on the mode of reproduction. Pooled samples are then subjected to high-throughput sequencing. Sequenced reads are subsequently aligned to the reference genome (shown at the bottom). We represent pool sequencing by an additional sampling step (called sampling step 2). We correct for both sampling steps when estimating N_e in pooled samples. Additionally, we take into account variable coverage levels across the genome (coverage R_{ij} for site i at $T = j, j \in \{0, t\}$) when correcting for the variance coming from sequencing.

t generations apart to estimate N_e (Figure 1) and denote the estimated effective population size by $\widehat{N}_e$. Multiallelic sites in populations with low mutation rates, such as *Drosophila*, exist but are rare and likely to be sequencing errors (Burke *et al.* 2010). Therefore we consider only biallelic SNPs at n polymorphic sites. At each site i ($i = 1 \ldots n$) the true population allele frequency is denoted by p_{ij} at time $T = j$, where $j \in \{0, t\}$. To obtain allele frequency estimates for an unknown p_{ij}, the population is subjected to sampling. We consider two sampling steps (Figure 1). At $T = j$, we first sample S_j individuals out of the population to create a pooled DNA library for sequencing. Note that the number of sampled individuals is constant over the n sites, and therefore the index i is omitted here. Sampling individuals can take place according to either plan I or plan II, as described above (also shown in Supplemental Material, Figure S1). As the second sampling step, we model Pool-seq by drawing R_{ij} reads out of the pooled DNA sample at each site i ($i = 1 \ldots n$). This allows for variation in sequence coverage. Below we derive the variance in allele frequency for a given site. To keep notation simple, we omit again the index i and denote the unknown sample allele frequency among the S_0 individuals at the first sampling time point ($T = 0$) by x and the subsequent allele frequency estimate obtained via pool sequencing from R_0 reads by $\hat{x}$. Similarly, at some $T = t$, the respective frequencies are denoted by y and $\hat{y}$. Note that under pool sequencing only $\hat{x}$ and $\hat{y}$ are observed.

Estimating N_e from temporal allele frequency changes

Under neutral Wright–Fisher evolution the variance in allele frequency (σ_p^2) generated by drift after t generations at a single locus in a diploid population is well described by the expression

$$\sigma_p^2 = p(1-p)\left[1 - \left(1 - \frac{1}{2N_e}\right)^t\right], \tag{1}$$

where p is the starting allele frequency (Falconer and Mackay 1996). Wright (1931) denoted the standardized variance by $F = \sigma_p^2/p(1-p)$, which leads to a convenient closed-form expression for N_e. Furthermore, if N_e is large enough, $F \approx 1 - e^{-t/2N_e}$ and N_e can be calculated as

$$N_e \approx \frac{-t}{2\ln(1-F)}. \tag{2}$$

The relation between N_e and allele frequency changes described in Equation 1 was first used by Krimbas and Tsakas (1971) in natural populations of olive flies. They estimated the variance using

$$F = F_a := \frac{1}{a}\sum_{k=1}^{a}\frac{(x_k - y_k)^2}{x_k(1 - x_k)}, \tag{3}$$

where x_k and y_k ($k = 1, \ldots, a$) are the observed allele frequencies in the samples collected t generations apart and a is the number of alleles at a specific locus. To eliminate the contribution of sampling errors to the variance, the total variance F_a was corrected for the random sampling noise by simply subtracting the corresponding variance. This approach was further investigated and developed by a number of authors (Pamilo and Varvio-Aho 1980; Nei and Tajima 1981; Pollak 1983; Waples 1989).

Possible sources of bias in N_e estimators were later investigated by Jorde and Ryman (2007). The authors pointed out that the expectation over F is typically approximated by taking the expected values separately for the numerator and the denominator (Turner *et al.* 2001). They suggested a different weighting scheme of alleles leading to an alternative less-biased estimator to measure temporal frequency change.

Correction for two-step sampling

We consider a random-mating population with discrete generations. Neutral evolution is assumed with no selection, migration, and mutation. Samples are drawn from the population at generations $T = 0$ and t. Throughout the derivation we consider diploid populations, and therefore a sample of S_j individuals leads to $2S_j$ sequences at times $T = j \in \{0, t\}$. Sampling is assumed to be binomial with parameters $2S_j$ and p_j (Waples 1989). In the second sampling step at time $T = j$, sequencing a random pool R_j of reads is also modeled as binomial sampling.

In analogy to Jorde and Ryman (2007), we use the following expression as our measure of the temporal change in allele frequency for biallelic sites,

$$F_c = \frac{(\hat{x}-\hat{y})^2}{\hat{z}-\widehat{xy}}, \qquad (4)$$

where $\hat{z} = (\hat{x}+\hat{y})/2$.

The expectation of F_c for a single biallelic locus is approximated by

$$E(F_c) \approx \frac{E(\hat{x}-\hat{y})^2}{E(\hat{z}-\widehat{xy})} = \frac{\text{Var}(\hat{x}) + \text{Var}(\hat{y}) - 2\text{Cov}(\hat{x},\hat{y})}{E(\hat{z}-\widehat{xy})}. \qquad (5)$$

For both plans, we derive expressions for the numerator and denominator in Equation 5 separately under the two-step sampling procedure, described above. Here we summarize our main conclusions; details on the derivation are provided in File S1. With $C_j := 1/2S_j + 1/R_j - 1/2S_jR_j$ for $j \in \{0, t\}$, and p denoting the true population allele frequency in the gamete pool at generation 0, we obtain

$$\text{Var}(\hat{x}) = p(1-p)C_0, \qquad (6)$$

and

$$\text{Var}(\hat{y}) = p(1-p)\left[1 - (1-C_t)\left(1 - \frac{1}{2N_e}\right)^t\right]. \qquad (7)$$

Note that Equations 6 and 7 differ only in the correction term C_j from that in Waples (1989).

Waples (1989) previously showed that the denominator in Equation 5 reduces to

$$E(\hat{z}-\widehat{xy}) = p(1-p) - \text{Cov}(\hat{x},\hat{y}). \qquad (8)$$

For plan II, $\text{Cov}(\hat{x},\hat{y}) = 0$ (Waples 1989), and F_c corrected for the noise coming from the two-step sampling for a single locus is given by

$$F_c' = \frac{F_c - C_0 - C_t}{1 - C_t}. \qquad (9)$$

For plan I, on the other hand, the sample allele frequency at generation 0 is positively correlated to the sample allele frequency at t because both are derived from the same population at generation 0. This requires us to calculate the sample covariance $\text{Cov}(\hat{x},\hat{y})$ in Equation 5. It turns out (see File S1 for details) that the covariance of $\hat{x}$ and $\hat{y}$ is equal to

$$\text{Cov}(\hat{x},\hat{y}) = \frac{p(1-p)}{2N}, \qquad (10)$$

where N is the census size of the population at generation 0. Equation 10 is in agreement with the corresponding term of the standard methods (Waples 1989). Substituting the inferred covariance into Equation 5 leads to the following corrected variance estimate, F_c' for plan I

$$F_c' = \frac{F_c(1 - 1/2N) - C_0 - C_t + 1/N}{1 - C_t}. \qquad (11)$$

We provide the corresponding formulas of F_c' in haploid populations in File S1.

With Pool-seq data, randomness in sequencing and local structures in the genome can lead to different coverage across marker sites, which we denote by R_{ij} for site i ($i = 1, \ldots n$) and time j ($j \in \{0, t\}$). In the genome-wide data set, we calculate F_c' across n SNPs by summing over all loci in the numerator and denominator separately before carrying out the division in Equation 9, leading to the following weighting scheme for plan II:

$$\overline{F_c'} = \frac{\sum_{i=1}^{n}(\hat{x}_i-\hat{y}_i)^2 - (\hat{z}_i - \hat{x}_i\hat{y}_i)(C_{i0} + C_{it})}{\sum_{i=1}^{n}(\hat{z}_i - \hat{x}_i\hat{y}_i)(1 - C_{it})}. \qquad (12)$$

Similarly, $\overline{F_c'}$ can be calculated for plan I using Equations 4 and 11. Analogous to the single-locus case, our proposed estimators $\widehat{N}_e(P)$ for a diploid population are obtained by plugging $\overline{F_c'}$ into Equation 2.

Long time series have recently become available for some E&R experiments (Barrick *et al.* 2009; Burke *et al.* 2010, 2014). Standard N_e estimators (Krimbas and Tsakas 1971; Nei and Tajima 1981; Waples 1989) assume a small number of generations (t) and approximate N_e using $2N_e \approx t/F$. If, however, t/N_e is larger, using this approximation can lead to severe bias (Figure S2). To avoid such a bias, we use Equation 2 to estimate N_e.

Simulations

We evaluate the performance of our estimator on data simulated under the neutral Wright–Fisher model. With a given population size of $2N_e$, we simulate the frequency trajectory of n independent SNPs. As we focus on biallelic SNPs, we assume two possible nucleotides (alleles) to be present in the population with given frequencies at the start. To create a new generation, nucleotides are drawn independently at random with a probability given by their respective allele frequencies in the old generation. The population is propagated at a constant size of $2N_e$ for t nonoverlapping generations. The effective population size is then estimated from allele frequencies inferred from Pool-seq samples taken from the population at the start and after t generations. The sampling of individuals to create the pooled DNA library is simulated by using sampling without replacement. To model the uneven coverage of genome-wide sequence data, we simulate a random coverage for each site, using a Poisson distribution with parameter equal to the given mean coverage. For every position, reads are then generated by binomial sampling from the library with sample size equal to the local coverage.

We assess the performance of our estimator for various combinations of N_e, pool size, coverage, number of SNPs, and distribution of starting allele frequencies. Additional to these parameters, we also test how the ratio between census and effective population size ($r = N/N_e$) affects the accuracy of the proposed estimator. For this purpose, we increment the population size to a desired level of N in each generation
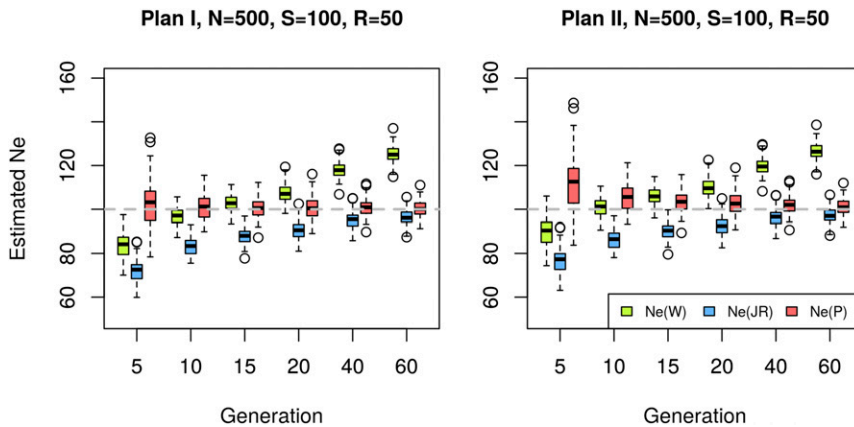
Figure 2 Effective population size estimated with different methods. Sixty generations of Wright–Fisher neutral evolution with $N_e = 100$ diploid individuals were simulated for $n = 2000$ unlinked loci (SNPs). Prior to sampling, the population was increased to a census size of $N = 500$ individuals at each generation. At the starting population and at each indicated time point a sample was taken to create a pool of $S = 100$ individuals. The pool was sequenced to an average coverage of $R = 50$ and N_e was estimated on the resulting data set by separately contrasting allele frequencies at generation 0 to each of the evolved generations denoted on the x-axis, using $N_e(P)$, $N_e(W)$ (Waples 1989), and $N_e(JR)$ (Jorde and Ryman 2007). Each box represents results from 100 simulations with identical parameters. The dashed gray line shows the true value of N_e. Data are simulated under plan I assumptions and the results of plan I and II estimators are shown in the left and right panels, respectively.

while keeping the allele frequencies unchanged to avoid introducing additional sampling variance. We simulated each scenario 100 times.

Linkage disequilibrium between loci can reduce the number of independent SNPs, thereby increasing the variance of the estimate. The impact of dependence between SNPs is investigated based on 10 replicated whole-genome forward simulations with recombination, using the software tool MimicrEE (Kofler and Schlötterer 2014). As a starting population for the forward simulations, we sampled 2000 haploid genomes out of 8000 genomes simulated with fastsimcoal v.1.1.2 (Excoffier and Foll 2011; Bastide *et al.* 2013). The parameters used to generate the genomes mimic a wild population of *D. melanogaster* from Vienna (Fiston-Lavier *et al.* 2010; Bastide *et al.* 2013; Kofler and Schlötterer 2014). Allele counts are subjected to binomial sampling to mimic Pool-seq with a given sequence coverage. N_e is estimated in nonoverlapping windows, each containing a fixed number of SNPs.

Estimating N_e on simulated data

We denote our estimator corrected for the additional sampling step, *i.e.*, pooling, by $N_e(P)$. We compare $N_e(P)$ to the standard estimators $N_e(W)$ and $N_e(JR)$ proposed by Waples (1989) and Jorde and Ryman (2007) that correct only for a single sampling step.

We illustrate experimental sampling procedures by considering two major scenarios: (i) The full population is sequenced as one large pool and (ii) only a subset of the population is used to create pooled samples. Under scenario (i) we simulate only a single binomial sampling step to represent sampling reads out of the DNA pool. The pool size is set to be equal to the census size of the population ($S_j = N$), and the number of sampled reads (R_{ij}) represents the per-site coverage. For estimators that correct only for a single sampling step, we use the coverage (R_{ij}) as the sample size. For scenario (ii), the sampled individuals (S_j) and the read number (R_{ij}) represent the pool size and coverage for $N_e(P)$. The coverage (R_{ij}) is taken as the sample size for the $N_e(W)$ and $N_e(JR)$ estimators, as these methods consider only one sampling step.

Change point inference for genome-wide estimates

The effect of genetic drift on the variance in allele frequency specified in Equation 1 holds only under the assumptions of Wright–Fisher neutral evolution. Deviations from the Wright–Fisher model, such as the presence of selection or demography, may cause systematically different changes in allele frequency, affecting the variance and causing locally variable patterns in genetic diversity. Furthermore, the effect of selection on one site of the genome may cause changes in the behavior of variants at nearby sites (Maynard Smith and Haigh 1974; Barton 2000; Comeron *et al.* 2008). As a result, the estimates of N_e at different locations of the genome may deviate from the true number of breeding individuals in the population (Kimura and Crow 1963; Charlesworth 2009). For example, regions under background selection are associated with reduced $\widehat{N}_e$ values that extend to linked sites due to the Hill–Robertson effect (Charlesworth 1996, 2012a; Comeron *et al.* 2008). Similarly, selectively favorable alleles can also drag nearby neutral sites to high frequency (Maynard Smith and Haigh 1974), causing a local reduction in the estimated N_e (Liu and Mittler 2008). Such an event is also known as a selective sweep (Berry *et al.* 1991). On the other hand, we expect the opposite pattern, *i.e.*, a local elevation of $\widehat{N}_e$ for types of selection such as balancing selection that maintain variation in the genome (Baysal *et al.* 2007; Charlesworth 2009).

To detect such patterns in $\widehat{N}_e$, we apply a segmentation algorithm to partition the genome into locally homogeneous $\widehat{N}_e$ stretches. We use a method related to an approach suggested by Futschik *et al.* (2014) for partitioning DNA sequences with respect to GC content. It is based on a statistical multiscale criterion and provides statistical error control, in the sense
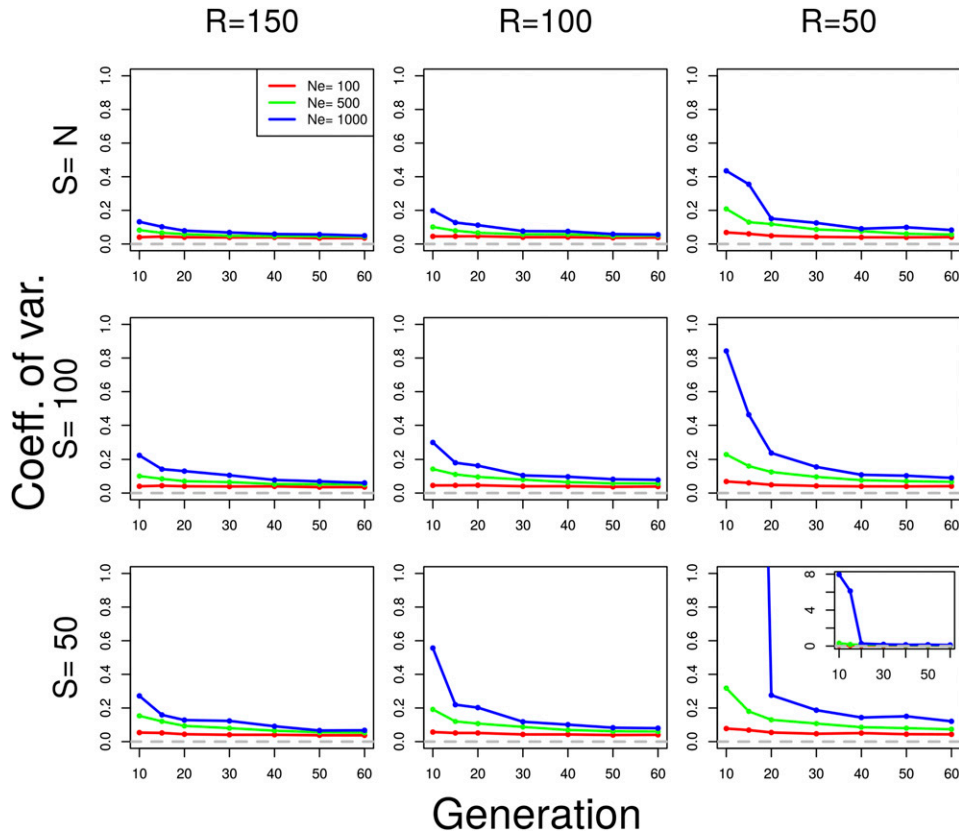
Figure 3 Coefficient of variation of $N_e(P)$ under plan I for various parameter values. Neutral Wright–Fisher simulations were performed with various combinations of the parameters: effective population size ($N_e = 100, 500, 1000$ diploid individuals), pool size ($S = 100, 50$), and coverage ($R = 150, 100, 50$). N_e was estimated with $N_e(P)$ under plan I, using $n = 2000$ SNPs. $S = N$ indicates scenarios when the whole population is sequenced as a single pool. For all simulations, we assumed $N = N_e$. Each value is calculated over 100 simulations. When the coefficient of variation exceeds one, the inset shows the actual value.

that the estimated number of windows will not exceed the true one except for a small error probability α to be specified by the user. With our N_e estimates, we use a criterion proposed by Frick *et al.* (2014) for normally distributed responses. It is implemented as part of the R package *stepR* (Frick *et al.* 2014). By using simulations with selection we also illustrate that this method is able to capture the signal of locally variable $\widehat{N}_e$ along the chromosome.

Data availability

We estimated N_e in an E&R study with *D. melaongaster*, published in Orozco-terWengel *et al.* (2012) and Franssen *et al.* (2015). Pool-seq read libraries from these studies are available at the European Sequence Read Archive at http://www.ebi.ac.uk/ena/ under accession nos. ERP001290 and ERS460611–ERS460613.

Results and Discussion

Two-step correction is vital to avoid large bias in $\widehat{N}_e$ with Pool-seq data

Methods that do not correct for the additional sampling step caused by pooling can lead to substantial bias in $\widehat{N}_e$ as illustrated in Figure 2. Using simulated data, we compare our proposed estimator $N_e(P)$ to two commonly used estimators $N_e(W)$ (Waples 1989) and $N_e(JR)$ (Jorde and Ryman 2007) that provide highly accurate estimates when only a single sampling event is simulated (Figure S3). Figure 2 shows that

the additional correction substantially decreases the bias for almost all scenarios (see also Figure S4, Figure S5, and Figure S6). Under plan I, $N_e(P)$ is nearly unbiased. The plan II version of the estimator has a slight upward bias when applied on data simulated under plan I, if the samples are taken at very close time points.

As an alternative approach, we also estimated N_e separately for each locus, using F_c' in Equations 9 and 11. We then calculated the effective population size across the n loci as the harmonic mean over the single-locus $\widehat{N}_e$ estimates ($\widehat{N}_e^*(P)$) (Peel *et al.* 2013). In our simulations, the harmonic mean estimator shows an accuracy similar to that of the original $\widehat{N}_e(P)$ (Figure S7). However, for t lying in the midrange of the simulated interval ($t = 15$–40), $\widehat{N}_e^*(P)$ is slightly more biased under plan I.

Because of the additional sampling variance, both $N_e(W)$ and $N_e(JR)$ have a downward bias in particular for small t. Furthermore, $N_e(W)$ is upwardly biased for larger values of t, probably reflecting that alleles closer to fixation or loss are contributing less to the variance (Waples 1989). The drift variance accumulates with an increasing number of generations, while the sampling variance stays constant, making the initial bias of $N_e(JR)$ less pronounced for larger t. When samples are collected only a few generations apart, the variance of $N_e(P)$ estimators tends to be larger than that of $N_e(W)$ and $N_e(JR)$ under both plans.

Plan I and II estimators differ by a factor resulting from the covariance between the sample frequencies at generations
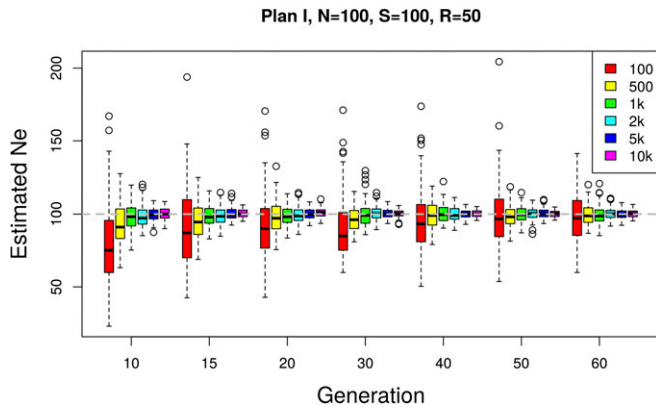
Figure 4 Effect of the number of SNPs used for estimating N_e. The effective population size is estimated using $N_e(P)$ plan I on simulated data with $N_e = N = 100$. A total number of $S = 100$ individuals are pooled and sequenced at a mean coverage of $R = 50$. Based on 100 simulation runs, N_e is estimated using different numbers of SNPs at multiple time points.
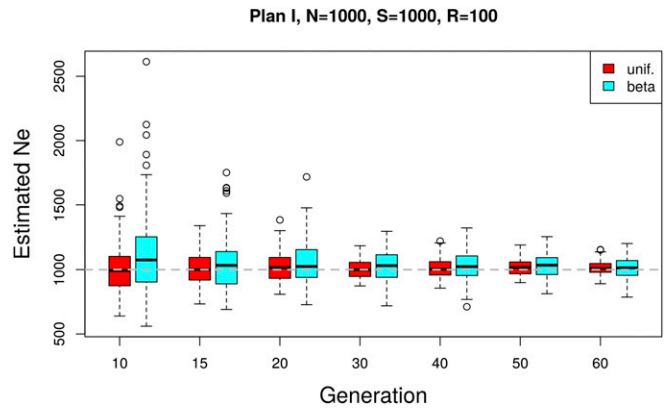

Figure 5 Influence of the starting allele frequency distribution on $N_e(P)$ under plan I. A comparison between uniform and Beta(0.2, 0.2)-distributed (neutral) starting allele frequencies is shown. The simulation parameters match those of the genome-wide simulations in Figure 6.

0 and t (Equation 10), which is inversely proportional to the census population size. Consequently, the difference between plans I and II becomes smaller for increasing N. Waples (1989) investigated how the ratio between census and effective population size ($r = N/N_e$) affects the accuracy of the estimators and concluded that the ratio of $r \geq 2$ is sufficient to reach similar estimates for both sampling schemes. We tested the performance of $N_e(P)$ on simulated data with $N_e = 100$ and $N : N_e$ ratios of $r = 1, 2, 5$ with different coverages and pool sizes (Figure S4, Figure S5, and Figure S6). When $N = N_e$, the $N_e(P)$ plan I method achieves highly accurate estimates for all time points in contrast to the other methods (Figure S4). If, however, the $N_e(P)$ plan II estimator is applied to data simulated under plan I, we observe an upward bias for small t, which improves with an increasing number of generations. This pattern is not unexpected since the missing covariance term becomes less influential in view of the increasing drift variance after several generations. When the entire population is sequenced as a single pool ($S = 100$), the plan II estimators of Waples (1989) and Jorde and Ryman (2007) perform similarly to the $N_e(P)$ plan I estimator because the correction for pooling in $N_e(P)$ cancels out the additional covariance term when $S = N$, making the term used as F approximately identical to that of $N_e(JR)$. This is a general pattern irrespective of r.

For $r \geq 2$, $N_e(P)$ plan I remains highly accurate (Figure S5 and Figure S6). Furthermore, when increasing the census size under a constant N_e (equivalent to increasing r), the covariance between sample allele frequencies decreases, making the difference between plans I and II almost negligible (Waples 1989). The sampling variance becomes proportionally smaller compared to the drift variance with an increasing number of generations between the samples. This improves our ability to accurately estimate N_e.

Correcting for the additional variance inherent to Pool-seq leads to an improved performance of $N_e(P)$ compared to the standard methods for both plans. In general, with Pool-seq data the extent of the bias of the $N_e(W)$ and $N_e(JR)$ estimates depends on the ratio between N and S, smaller sample sizes (S) leading to a larger bias. As we accounted for the sequencing step with these estimators (*Estimating N_e on simulated data*), decreasing the coverage at a given pool size does not change the bias much but rather increases the variance of the estimators.

In most of the experimental studies the investigator has control over the census population size; thus requiring the knowledge of N for $N_e(P)$ plan I does not restrict the analysis. We illustrate the performance of $N_e(P)$ plan I only when $N_e = N$ in the main text, but according to Figure S5 and Figure S6, $N_e(P)$ plan I is also highly accurate when $r \geq 2$.

We show the coefficient of variation (CV) of the $N_e(P)$ plan I estimator in Figure 3. The CV is defined as the ratio between the standard deviation and the mean (CV $= \hat{\sigma}/\hat{\mu}$, where both $\hat{\sigma}$ and $\hat{\mu}$ are estimated from the sample). It measures the relative dispersion of the distribution of the estimated values. $N_e(P)$ estimators are highly precise in nearly all cases, except when the drift variance is negligible compared to the sampling variance (Figure 3; see also Figure S9 and Figure S11 where $N_e = 1000$, $t < 30$, $S \leq 100$, and $R = 50$). The bias is coming from a few outlier estimates, but the median shows more robust results (Figure S13). For plan II estimators, the behavior of the method is similar (Figure S8, Figure S10, Figure S12, and Figure S14). Note that the simulations underlying Figure S8, Figure S10, Figure S12, and Figure S14 have been done under plan I.

Increasing the number of SNPs reduces the variance of $N_e(P)$

We test how the number of loci (n) used to infer N_e affects the accuracy and the precision of the estimates by gradually increasing the number of independent SNPs from 100 to 10,000 (Figure 4). We observe a larger variance and a slight downward bias for a small number of SNPs (100 SNPs). Both the bias and the variance become smaller with a larger
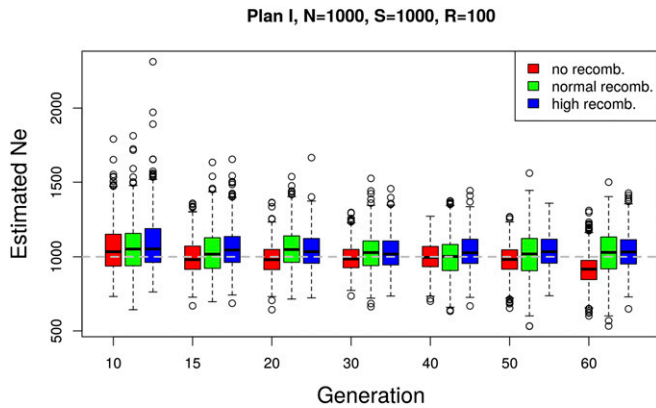
Figure 6 Effect of linkage disequilibrium on $\widehat{N}_e$. The effect of linkage disequilibrium on our estimator was evaluated based on a whole-genome forward simulation with recombination using the software MimicrEE (Kofler and Schlötterer 2014). Three sets of simulations were performed with different rates of recombination: high, normal, and no recombination. For each parameter setup, a genome-wide simulation is replicated 10 times. The effective population size was estimated with $N_e(P)$ (plan I) in nonoverlapping windows of $n = 10,000$ SNPs for each replicate. The box plots show the distribution of N_e estimates across replicates and windows.

number of SNPs. Some further improvement is obtained when >10,000 SNPs are used (not shown), but the benefit of additional independent SNPs levels off. We conclude that $n = 2000$ SNPs usually provide sufficient precision and accuracy. However, when linkage disequilibrium is present in a genome-wide data set, the number of truly independent SNPs per window is reduced and a larger number of loci is recommended.

A skewed starting allele frequency distribution only moderately increases the variance of $N_e(P)$

In natural populations, the neutral site frequency spectrum is skewed toward allele frequencies close to the boundaries. $N_e(P)$ uses a weighting scheme that is not very sensitive to this skew (see also Jorde and Ryman 2007). This makes it robust with respect to the shape of the starting allele frequency distribution. We illustrate this with a simulated data set having a starting allele frequency distribution that is skewed toward low- and high-frequency variants (Beta(0.2, 0.2)) as expected under neutrality. The estimates of N_e from such data sets are compared to simulated data with matching parameters but uniform starting allele frequency distribution (Figure 5). We observe a very slight upward bias with neutral starting allele frequencies compared to uniform and a moderate increase in the variance given $t \geq 15$. With an increasing number of generations, the difference becomes negligible.

The presence of linkage disequilibrium does not have a large effect on the precision of $N_e(P)$

We investigated the sensitivity of our estimator to linkage disequilibrium between loci, using genome-wide neutral simulations with recombination (Kofler and Schlötterer 2014).

We simulated data with three different rates of recombination: high, normal, and no recombination. For the first case, the recombination rate is set to mimic the behavior of almost independent SNPs. In the normal recombination rate scenario, we use *D. melanogaster* recombination rates (Fiston-Lavier *et al.* 2010). The effective population size was estimated in nonoverlapping windows with a fixed number of $n = 10,000$ SNPs (Figure 6). Different levels of linkage disequilibrium affect the number of independent loci per window. Nevertheless, we observe only a slight increase in the precision of the N_e estimates with increasing recombination rate (Figure 6).

Heterogeneous $\widehat{N}_e$ along the genome in an E&R study with *D. melanogaster*

We estimated N_e in a recent E&R study with *D. melanogaster* (Orozco-terWengel *et al.* 2012; Franssen *et al.* 2015). In this experiment replicate populations of 1000 individuals were subjected to a fluctuating hot environment for 59 generations. Allele frequency estimates were obtained for founder and evolved populations, using Pool-seq. N_e was estimated based on the allele frequency changes between founder and latest evolved populations in nonoverlapping windows containing 10,000 SNPs, using $N_e(P)$ under plan I. To determine the number of DNA stretches with different $\widehat{N}_e$ along the genome, we use a segmentation algorithm provided in the software tool by Frick *et al.* (2014). This method requires homogeneity of variances. Since the variance of estimates increases with N_e, the estimates were log-transformed before applying the partitioning procedure. The obtained step function was back-transformed to the original scale and is shown for three biological replicates (Figure 7).

The mean and the median estimates for each chromosome arm as well as across the genome are stable across replicates (see Table 1 and Table S1). As experimental evolution studies often aim to find signals that are consistent across replicates, this can be an important check of the experimental setup. On the other hand, we see differences between chromosome arms. For example, the mean is clearly lower for 3R, emphasizing the added value of spatial analysis compared to genome-wide estimates.

In *D. melanogaster* $\widehat{N}_e$ ranges between ~100 and 400. Around the centromere of chromosome 2, the estimated N_e decreases by two-thirds in replicates 1 and 3, which is in agreement with the expectation of low diversity and, as a consequence, low N_e in regions with reduced recombination (Begun and Aquadro 1992; Presgraves 2005; Haddrill *et al.* 2007; Campos *et al.* 2012). Furthermore, $\widehat{N}_e$ is low on the entire chromosome arm 3R and also on parts of 3L. Overall, these patterns can be attributed to strong LD, caused either by low recombination rates around the centromeres (Chan *et al.* 2012) or by segregating inversions (Kapun *et al.* 2014) in combination with selection potentially on rare variants. The reduction in $\widehat{N}_e$ is also well captured by the segmentation algorithm (Figure 7), which shows a similar pattern when applied on simulated data with selection (Figure S15). These
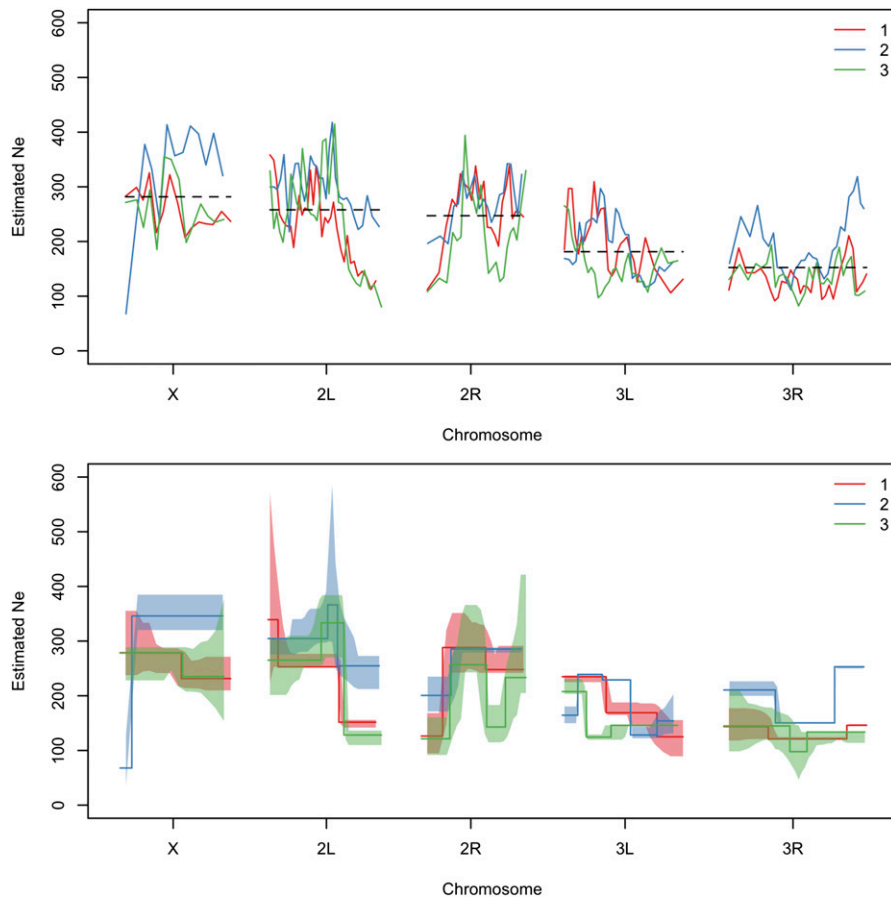
Figure 7 Genome-wide $\widehat{N}_e$ from an E&R study with *D. melanogaster*. N_e is estimated based on the allele frequency changes between founder and evolved populations at generation 59 (Franssen *et al.* 2015). In the top panel, we show genome-wide estimates calculated with $N_e(P)$ (plan I), using $N = 1000$ as census size and $S = 500$ as pool size (Orozco-terWengel *et al.* 2012) and non-overlapping windows of 10,000 SNPs. Chromosome-wide mean estimates across replicates are shown by the dashed lines and also calculated separately for each replicate in Table 1. DNA stretches with significantly different $\widehat{N}_e$ are determined using the *stepR* software package (Frick *et al.* 2014) (bottom panel). Lower and upper $1 - \alpha$ confidence bands are shown as shaded areas. α controls the error, *i.e.*, the probability for overestimating the number of change points, and is calculated automatically as described in Frick *et al.* (2014). The colors indicate different biological replicates.

results are consistent with those of Tobler *et al.* (2014), who observed a massive amount of outlier SNPs around the centromere of chromosome 2 and on 3R. Interestingly, certain regions of the genome show extensive differences in $\widehat{N}_e$ between the replicates, which might be reflecting different selection histories or differences in demography, such as replicate-specific bottlenecks.

$\widehat{N}_e$ may also vary as a result of differences in the modes of transmission of different components in the genome. For example, on the X chromosome, N_e is equal to three-quarters of the autosomal population size (Vicoso and Charlesworth 2006, 2009). Interestingly, our estimates in the E&R experiment do not reflect this expectation of reduced effective population size. Instead, we estimate N_e to be as high as $\widehat{N}_e$ on the autosomes. Unequal sex ratio between males and females can be a source of such a pattern (Charlesworth 2009); however, unbalanced sex ratio has not been reported in this experiment. Another possible explanation for increased $\widehat{N}_e$ on the X can be the presence of background selection as suggested by Charlesworth (2012b). He argues that because of the lack of recombination in male *Drosophila*, the effect of background selection is more effective on the autosomes than on the X chromosome. Orozco-terWengel *et al.* (2012) reported differences in the number of putatively selected sites between the X and autosomes. They found that candidate SNPs were underrepresented on the X. Their selection scan

identifies signatures of deviation from neutral expectation, which is also reflected in the reduction in $\widehat{N}_e$ on the autosomes, indicating higher selection pressure.

Recommendations for genome-wide data sets

Most of the methods proposed previously are not designed for genome-wide high-density SNP data sets. However, the method of Jorde and Ryman (2007) was successfully used for genome-wide data by Foll *et al.* (2014). Reed *et al.* (2014) also used a similar approach to estimate N_e for whole-genome data, using sliding windows. We estimated N_e in windows with a fixed number of SNPs. Using windows of fixed lengths in base pairs would affect the variance of the estimator (Figure 4) but does not distort the mean. All these approaches, however, do not account for the ruggedness of the recombination landscape and can lead to windows with different levels of linkage disequilibrium in them. To overcome this problem it would be possible to define windows based on recombination distance. Unfortunately, the lack of haplotype information in the Pool-seq data makes it difficult to infer linkage disequilibrium. One way to infer linkage information from pooled sequence data is provided by the software LDx (Feder *et al.* 2012). For model organisms such as *Drosophila*, readily available recombination maps can also be used as a proxy (Przeworski *et al.* 2001; Kulathinal *et al.* 2008; Fiston-Lavier *et al.* 2010). If only a single genome-wide N_e estimate

Table 1 Genome-wide mean $\widehat{N}_{\mathrm{e}}$ from an E&R study with _D. melanogaster_

	Mean					
Replicate	X	2L	2R	3L	3R	Genome-wide
R1	257.9675	231.6854	257.0828	193.4339	131.7072	199.4463
R2	328.8878	297.9832	274.8529	193.3237	194.9571	239.3618
R3	263.4829	246.5448	211.8995	157.6411	133.9459	187.1573

The effective population size is estimated with $N_{\mathrm{e}}(P)$ plan I in windows of 10,000 SNPs (Figure 7). The mean estimates across windows are shown for the major chromosome arms. The genome-wide mean is taken over the autosomes, excluding chromosome 4.

is required, one can alternatively use a set of randomly distributed SNPs over the genome to obtain $\widehat{N}_{\mathrm{e}}$.

Temporal methods make a number of assumptions, which, if violated, can lead to biased N_{e} estimates. For example, in our simulations, we considered only effective population sizes that are constant over time. Fluctuating N_{e} is a frequent phenomenon in natural populations and can be an important component of an experimental design. For example, in repeatedly bottlenecked populations, the smallest population size dominates $\widehat{N}_{\mathrm{e}}$ (Luikart _et al._ 1999; Charlesworth 2009). But even in strictly controlled populations the experimental regime can induce changes in N_{e}. When the population changes in size, the estimated N_{e} is generally interpreted as the harmonic mean of the effective population sizes over the generations (Wright 1938; Nei and Tajima 1981; Waples 1989). However, if time series allele frequency data are available, such changes can be detected by estimating N_{e} from pairwise comparisons between consecutive time points.

All evolutionary forces (selection, demography, etc.) that lead to deviations from the neutral expectation will also affect our estimate. Nevertheless, systematic forces that result in locally different values of $\widehat{N}_{\mathrm{e}}$ can be detected with a sliding-window approach, as illustrated with simulations under selection (Figure S15). The _D. melanogaster_ data set also illustrates this point; _i.e._, the hypothesized regions under selection coincide with regions of reduced N_{e} (Orozco-terWengel _et al._ 2012; Tobler _et al._ 2014; Franssen _et al._ 2015). For this to be detected, however, most of the allele frequency change has to occur over the sampled time span.

In the E&R study with _D. melanogaster_, shown above, the criterion of nonoverlapping generations, assumed by temporal methods, is met (see Orozco-terWengel _et al._ 2012 for details on experimental design). However, for samples from an age-structured population, the resulting $\widehat{N}_{\mathrm{e}}$ can be biased (Waples and Yokota 2007). In these cases, as suggested by Waples and Yokota (2007), larger spacing between samples maximizes the drift signal compared to sampling biases associated with age structure.

Using a small number of generations can lead to outlier estimates

In general, $N_{\mathrm{e}}(P)$ has a lower bias but larger variance, especially when t is small. As pointed out by Jorde and Ryman (2007) our weighting scheme leads to an increased variance but a smaller bias compared to other schemes. We observe outlier estimates among replicates at early generations (gen-

eration 5, Figure 2, Figure S4, Figure S5, and Figure S6) for $N_{\mathrm{e}}(P)$. When the sampling variance is large compared to the drift variance ($N_{\mathrm{e}} = 1000$, $S \leq 100$, and $R = 50$, Figure S11 and Figure S12), the deviation of the outlier estimates from the true N_{e} is particularly large. For a few cases, we even observe large negative estimates. Negative estimates, in general, can be interpreted as N_{e} being infinity, that is, no evidence of genetic drift (Peel _et al._ 2013). In our simulations this is plausible when N_{e} is large and t is small, such that drift has not had a large effect on the population allele frequencies yet. Note that the harmonic mean estimator ($\widehat{N}_{\mathrm{e}}^{*}(P)$) has smaller variance for large N_{e} (Figure S16). This estimator, however, is less accurate than $N_{\mathrm{e}}(P)$ for small N_{e} as shown in Figure S17.

To eliminate potential outliers and an inflated variance we recommend increasing the signal-to-noise ratio by pooling a sufficient number of individuals. Using later generations or increasing the number of SNPs in the analysis also helps to avoid outlier estimates. When none of these strategies can be applied, we suggest using the genome-wide median $N_{\mathrm{e}}(P)$ estimates or the harmonic mean estimator, as these are more robust to extreme outliers.

Conclusions

Effective population size is an important parameter for describing evolutionary dynamics, making its accurate estimation essential for population genetic studies. Several methods have been designed to estimate N_{e}, and their performance was comprehensively evaluated on simulated as well as real data (Barker 2011; Serbezov _et al._ 2012; Baalsrud _et al._ 2014; Holleley _et al._ 2014; Gilbert and Whitlock 2015). These studies mainly focused on genetic data collected from natural populations, which usually differ from experimental studies in terms of the census population size and sampling scheme. We designed a method that accurately infers the effective population size in genome-wide data from experimental populations sequenced in pools. Our approach improves temporal methods by explicitly correcting for two stages of sampling introduced by pooling and sequencing. Our results on simulated data confirm that methods that fail to properly account for the two stages of sampling inherent to Pool-seq can lead to severely biased N_{e} estimates.

Pool-seq data are often considered to be overdispersed, _i.e._, displaying more variability than is predicted by the binomial sampling model (Yang _et al._ 2012). However, Zhu _et al._ (2012) and Futschik and Schlötterer (2010) validated that

the error in allele frequency estimates is reasonably well approximated by binomial sampling given that a large enough number of individuals are pooled. Nevertheless, if overdispersion is present in the data, that will lead to additional variance, which is not modeled in our framework and will result in a downward bias of the estimated N_e. If the level of overdispersion can be inferred for the data (see, *e.g.*, Gautier *et al.* 2013; Illingworth 2015), it is possible to introduce a parameter that accounts for the additional between-pool variation (see File S1, Equation S8).

We also illustrate the applicability of our method for estimating N_e from experimental data of *D. melanogaster* and show that in combination with a recursive partitioning method we can infer patterns of local variation in N_e along the genome. Additionally, it is possible to calculate confidence intervals based on the χ^2 distribution (Waples 1989) or alternatively apply a nonparametric bootstrap approach.

Software availability

Our proposed estimators along with standard methods from the literature are implemented within the R package *Nest*. The package is currently available at https://github.com/ThomasTaus/Nest.

Acknowledgments

Literature Cited

Anderson, E. C., E. G. Williamson, and E. A. Thompson, 2000 Monte Carlo evaluation of the likelihood for N(e) from temporally spaced samples. Genetics 156(4): 2109–2118.

Baalsrud, H. T., B.-E. Saether, I. J. Hagen, A. M. Myhre, T. H. Ringsby *et al.*, 2014 Effects of population characteristics and structure on estimates of effective population size in a house sparrow metapopulation. Mol. Ecol. 23(11): 2653–2668.

Barker, J. S. F., 2011 Effective population size of natural populations of *Drosophila buzzatii*, with a comparative evaluation of nine methods of estimation. Mol. Ecol. 20(21): 4452–4471.

Barrick, J. E., D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh *et al.*, 2009 Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature 461(7268): 1243–1247.

Barton, N. H., 2000 Genetic hitchhiking. Philos. Trans. R. Soc. Lond. B Biol. Sci. 355(1403): 1553–1562.

Bastide, H., A. Betancourt, V. Nolte, R. Tobler, P. Stöbe *et al.*, 2013 A genome-wide, fine-scale map of natural pigmentation variation in *Drosophila melanogaster*. PLoS Genet. 9(6): e1003534.

Baysal, B. E., E. C. Lawrence, and R. E. Ferrell, 2007 Sequence variation in human succinate dehydrogenase genes: evidence for long-term balancing selection on SDHA. BMC Biol. 5: 12.

Begun, D. J., and C. F. Aquadro, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature 356(6369): 519–520.

Berry, A. J., J. W. Ajioka, and M. Kreitman, 1991 Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. Genetics 129: 1111–1117.

Boitard, S., R. Kofler, P. Françoise, D. Robelin, C. Schlötterer *et al.*, 2013 Pool-hmm: a Python program for estimating the allele frequency spectrum and detecting selective sweeps from next generation sequencing of pooled samples. Mol. Ecol. Resour. 13 (2): 337–340.

Burke, M. K., J. P. Dunham, P. Shahrestani, K. R. Thornton, M. R. Rose *et al.*, 2010 Genome-wide analysis of a long-term evolution experiment with *Drosophila*. Nature 467(7315): 587–590.

Burke, M. K., G. Liti, and A. D. Long, 2014 Standing genetic variation drives repeatable experimental evolution in outcrossing populations of *Saccharomyces cerevisiae*. Mol. Biol. Evol. 31(12): 3228–3239.

Campos, J. L., B. Charlesworth, and P. R. Haddrill, 2012 Molecular evolution in nonrecombining regions of the *Drosophila melanogaster* genome. Genome Biol. Evol. 4(3): 278–288.

Chan, A. H., P. A. Jenkins, and Y. S. Song, 2012 Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. PLoS Genet. 8(12): e1003090.

Charlesworth, B., 1996 Background selection and patterns of genetic diversity in *Drosophila melanogaster*. Genet. Res. 68(2): 131–149.

Charlesworth, B., 2009 Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. 10(3): 195–205.

Charlesworth, B., 2012a The effects of deleterious mutations on evolution at linked sites. Genetics 190(1): 5–22.

Charlesworth, B., 2012b The role of background selection in shaping patterns of molecular evolution and variation: evidence from variability on the *Drosophila* X chromosome. Genetics 191: 233–246.

Comeron, J. M., A. Williford, and R. M. Kliman, 2008 The Hill-Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. Heredity 100(1): 19–31.

Excoffier, L., and M. Foll, 2011 Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. Bioinformatics 27(9): 1332–1334.

Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to Quantitative Genetics*. Benjamin-Cummings, Menlo Park, CA.

Feder, A. F., D. A. Petrov, and A. O. Bergland, 2012 LDx: estimation of linkage disequilibrium from high-throughput pooled resequencing data. PLoS One 7(11): e48588.

Ferretti, L., S. E. Ramos-Onsins, and M. Pérez-Enciso, 2013 Population genomics from pool sequencing. Mol. Ecol. 22(22): 5561–5576.

Fisher, R., 1930 *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford.

Fiston-Lavier, A.-S., N. D. Singh, M. Lipatov, and D. A. Petrov, 2010 *Drosophila melanogaster* recombination rate calculator. Gene 463(1–2): 18–20.

Foll, M., Y.-P. Poh, N. Renzette, A. Ferrer-Admetlla, C. Bank *et al.*, 2014 Influenza virus drug resistance: a time-sampled population genetics perspective. PLoS Genet. 10(2): e1004185.

Foll, M., H. Shim, and J. D. Jensen, 2015 WFABC: a Wright-Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. Mol. Ecol. Resour. 15(1): 87–98.

Franssen, S. U., V. Nolte, R. Tobler, and C. Schlötterer, 2015 Patterns of linkage disequilibrium and long range hitchhiking in

evolving experimental *Drosophila melanogaster* populations. Mol. Biol. Evol. 32(2): 495–509.

Frick, K., A. Munk, and H. Sieling, 2014   Multiscale change point inference. J. R. Stat. Soc. Ser. B Stat. Methodol. 76(3): 495–580.

Futschik, A., and C. Schlötterer, 2010   The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. Genetics 186(1): 207–218.

Futschik, A., T. Hotz, A. Munk, and H. Sieling, 2014   Multiscale DNA partitioning: statistical evidence for segments. Bioinformatics 30(16): 2255–2262.

Gautier, M., J. Foucaud, K. Gharbi, T. Cézard, M. Galan *et al.*, 2013   Estimation of population allele frequencies from next-generation sequencing data: pool-*vs.* individual-based genotyping. Mol. Ecol. 22(14): 3766–3779.

Gilbert, K. J., and M. C. Whitlock, 2015   Evaluating methods for estimating local effective population size with and without migration. Evolution 69(8): 2154–2166.

Haddrill, P. R., D. L. Halligan, D. Tomaras, and B. Charlesworth, 2007   Reduced efficacy of selection in regions of the *Drosophila* genome that lack crossing over. Genome Biol. 8(2): R18.

Hill, W. G., 1981   Estimation of effective population size from data on linkage disequilibrium. Genet. Res. 38: 209–216.

Holleley, C. E., R. A. Nichols, M. R. Whitehead, A. T. Adamack, M. R. Gunn *et al.*, 2014   Testing single-sample estimators of effective population size in genetically structured populations. Conserv. Genet. 15: 23–35.

Huang, Y., S. I. Wright, and A. F. Agrawal, 2014   Genome-wide patterns of genetic variation within and among alternative selective regimes. PLoS Genet. 10(8): e1004527.

Hui, T.-Y. J., and A. Burt, 2015   Estimating effective population size from temporally spaced samples with a novel, efficient maximum-likelihood algorithm. Genetics 200: 285–293.

Illingworth, C. J. R., 2015   Fitness inference from short-read data: within-host evolution of a reassortant H5N1 influenza virus. Mol. Biol. Evol. 32(11): 3012–3026.

Jorde, P. E., and N. Ryman, 2007   Unbiased estimator for genetic drift and effective population size. Genetics 177: 927–935.

Kapun, M., H. van Schalkwyk, B. McAllister, T. Flatt, and C. Schlötterer, 2014   Inference of chromosomal inversion dynamics from Pool-Seq data in natural and laboratory populations of *Drosophila melanogaster*. Mol. Ecol. 23(7): 1813–1827.

Kawecki, T. J., R. E. Lenski, D. Ebert, B. Hollis, I. Olivieri *et al.*, 2012   Experimental evolution. Trends Ecol. Evol. 27(10): 547–560.

Kimura, M., 1964   Diffusion model in population genetics. J. Appl. Probab. 1: 177–223.

Kimura, M., and J. F. Crow, 1963   The measurement of effective population number. Evolution 17(3): 279–288.

Kofler, R., and C. Schlötterer, 2014   A guide for the design of evolve and resequencing studies. Mol. Biol. Evol. 31(2): 474–483.

Kofler, R., P. Orozco-terWengel, N. De Maio, R. V. Pandey, V. Nolte *et al.*, 2011a   PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. PLoS One 6(1): e15925.

Kofler, R., R. V. Pandey, and C. Schlötterer, 2011b   PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics 27(24): 3435–3436.

Kolaczkowski, B., A. D. Kern, A. K. Holloway, and D. J. Begun, 2011   Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. Genetics 187: 245–260.

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

Kulathinal, R. J., S. M. Bennett, C. L. Fitzpatrick, and M. A. F. Noor, 2008   Fine-scale mapping of recombination rate in *Drosophila*

refines its correlation to diversity and divergence. Proc. Natl. Acad. Sci. USA 105(29): 10051–10056.

Liu, Y., and J. E. Mittler, 2008   Selection dramatically reduces effective population size in HIV-1 infection. BMC Evol. Biol. 8: 133.

Long, A., G. Liti, A. Luptak, and O. Tenaillon, 2015   Elucidating the molecular architecture of adaptation via evolve and resequence experiments. Nat. Rev. Genet. 16(10): 567–582.

Luikart, G., J. M. Cornuet, and F. W. Allendorf, 1999   Temporal changes in allele frequencies provide estimates of population bottleneck size. Conserv. Biol. 13(3): 523–530.

Maynard Smith, J., and J. Haigh, 1974   The hitch-hiking effect of a favourable gene. Genet. Res. 23: 23–35.

Nei, M., and F. Tajima, 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. Evol. Appl. 1(3): 462–474.

Orozco-terWengel, P., M. Kapun, V. Nolte, R. Kofler, T. Flatt *et al.*, 2012   Adaptation of *Drosophila* to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. Mol. Ecol. 21(20): 4931–4941.

Pamilo, P., and S. L. Varvio-Aho, 1980   On the estimation of population size from allele frequency changes. Genetics 95: 1055–1057.

Peel, D., R. S. Waples, G. M. Macbeth, C. Do, and J. R. Ovenden, 2013   Accounting for missing data in the estimation of contemporary genetic effective population size (N(e)). Mol. Ecol. Resour. 13(2): 243–253.

Pollak, E., 1983   A new method for estimating the effective population size from allele frequency changes. Genetics 104: 531–548.

Presgraves, D. C., 2005   Recombination enhances protein adaptation in *Drosophila melanogaster*. Curr. Biol. 15(18): 1651–1656.

Przeworski, M., J. D. Wall, and P. Andolfatto, 2001   Recombination and the frequency spectrum in *Drosophila melanogaster* and *Drosophila simulans*. Mol. Biol. Evol. 18(3): 291–298.

Pudovkin, A. I., D. V. Zaykin, and D. Hedgecock, 1996   On the potential for estimating the effective number of breeders from heterozygote-excess in progeny. Genetics 144: 383–387.

Reed, L. K., K. Lee, Z. Zhang, L. Rashid, A. Poe *et al.*, 2014   Systems genomics of metabolic phenotypes in wild-type *Drosophila melanogaster*. Genetics 197: 781–793.

Schlötterer, C., R. Tobler, R. Kofler, and V. Nolte, 2014   Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. Nat. Rev. Genet. 15(11): 749–763.

Schlötterer, C., R. Kofler, E. Versace, R. Tobler, and S. U. Franssen, 2015   Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. Heredity 114(5): 431–440.

Serbezov, D., P. E. Jorde, L. Bernatchez, E. M. Olsen, and L. A. Vllestad, 2012   Short-term genetic changes: evaluating effective population size estimates in a comprehensively described brown trout (*Salmo trutta*) population. Genetics 191: 579–592.

Tallmon, D. A., A. Koyuk, G. Luikart, and M. A. Beaumont, 2008   Computer Programs: onesamp: a program to estimate effective population size using approximate Bayesian computation. Mol. Ecol. Resour. 8(2): 299–301.

Tobler, R., S. U. Franssen, R. Kofler, P. Orozco-terWengel, V. Nolte *et al.*, 2014   Massive habitat-specific genomic response in *D. melanogaster* populations during experimental evolution in hot and cold environments. Mol. Biol. Evol. 31(2): 364–375.

Turner, T. L., and P. M. Miller, 2012   Investigating natural variation in *Drosophila* courtship song by the evolve and resequence approach. Genetics 191: 633–642.

Turner, T. F., L. A. Salter, and J. R. Gold, 2001   Temporal-method estimates of N_e from highly polymorphic loci. Conserv. Genet. 2: 297–308.

Turner, T. L., A. D. Stewart, A. T. Fields, W. R. Rice, and A. M. Tarone, 2011 Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in *Drosophila melanogaster*. PLoS Genet. 7(3): e1001336.

Vicoso, B., and B. Charlesworth, 2006 Evolution on the X chromosome: unusual patterns and processes. Nat. Rev. Genet. 7: 645–653.

Vicoso, B., and B. Charlesworth, 2009 Effective population size and the faster-X effect: an extended model. Evolution 63(9): 2413–2426.

Wang, J., 2001 A pseudo-likelihood method for estimating effective population size from temporally spaced samples. Genet. Res. 78(3): 243–257.

Wang, J., 2009 A new method for estimating effective population sizes from a single sample of multilocus genotypes. Mol. Ecol. 18(10): 2148–2164.

Wang, J., 2013 A simulation module in the computer program COLONY for sibship and parentage analysis. Mol. Ecol. Resour. 13(4): 734–739.

Waples, R. S., 1989 A generalized approach for estimating effective population size from temporal changes in allele frequency. Genetics 121: 379–391.

Waples, R. S., and C. Do, 2008 LDNe: a program for estimating effective population size from data on linkage disequilibrium. Mol. Ecol. Resour. 8(4): 753–756.

Waples, R. S., and C. Do, 2010 Linkage disequilibrium estimates of contemporary N_e using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. Evol. Appl. 3(3): 244–262.

Waples, R. S., and P. R. England, 2011 Estimating contemporary effective population size on the basis of linkage disequilibrium in the face of migration. Genetics 189: 633–644.

Waples, R. S., and M. Yokota, 2007 Temporal estimates of effective population size in species with overlapping generations. Genetics 175: 219–233.

Williamson, E. G., and M. Slatkin, 1999 Using maximum likelihood to estimate population size from temporal changes in allele frequencies. Genetics 152: 755–761.

Wright, S., 1931 Evolution in Mendelian populations. Genetics 16: 97–159.

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

Yang, X., J. A. Todd, D. Clayton, and C. Wallace, 2012 Extrabinomial variation approach for analysis of pooled DNA sequencing data. Bioinformatics 28(22): 2898–2904.

Zhu, Y., A. O. Bergland, J. González, and D. A. Petrov, 2012 Empirical validation of pooled whole genome population re-sequencing in *Drosophila melanogaster*. PLoS One 7(7): e41901.

Communicating editor: M. A. Beaumont

Supplement

Ágnes Jónás, Thomas Taus, Carolin Kosiol, Christian Schlötterer, Andreas Futschik

1 Normalized allele frequency changes under two-step sampling

1.1 Calculating the expected normalized variation

In the following, we calculate the expected value for a measure of allele frequency change due to the combination of t generations of drift and sampling. As in Waples (1989), we consider

$$F_c = \frac{\sum_{i=1}^{a}(\hat{x}_k - \hat{y}_k)^2}{\sum_{k=1}^{a} \hat{z}_k - \hat{x}_k\hat{y}_k} \ , \tag{S1}$$

as our measure, where $\hat{x}$ and $\hat{y}$ denote the observed allele frequencies in the pooled DNA sample collected t generations apart, a is the number of alleles at a locus and $\hat{z}_k := (\hat{x}_k + \hat{y}_k)/2$. We consider only biallelic sites, which reduces equation (S1) to

$$F_c = \frac{(\hat{x} - \hat{y})^2}{\hat{z} - \hat{x}\hat{y}} \tag{S2}$$

for a given locus. The expectation of F_c is approximated in the following way:

$$E(F_c) \approx \frac{E(\hat{x} - \hat{y})^2}{E(\hat{z} - \hat{x}\hat{y})} = \frac{Var(\hat{x}) + Var(\hat{y}) - 2Cov(\hat{x}, \hat{y})}{E(\hat{z} - \hat{x}\hat{y})}. \tag{S3}$$

We will provide expressions for the numerator and denominator in equation (S3) both for plan I and II under the two-step sampling process. For this purpose, we introduce the following notation: Let p and p_t denote the true allele frequency in the gamete pool preceding the first sampled generation (0) and after t generations, respectively. S_0 is the number of individuals sampled at the first time point (generation 0), and x denotes the relative allele frequency in this sample. As in the main text, we distinguish between the allele frequencies (x, y) estimated after sampling individuals at generations 0 and t, and the frequencies $(\hat{x}, \hat{y})$ after pool sequencing based on R_j ($j \in \{0, t\}$) reads. For a visual representation of our sampling schemes along with the notation see Fig. S1.

We refer to previous findings of Waples (1989) without carrying out the detailed derivations here, namely

1. The variance of x is binomial under both plans:

$$Var(x) = E(x - p)^2 = \frac{p(1 - p)}{2S_0}. \tag{S4}$$

2. The variance due to t generations of genetic drift is:

$$Var(p_t|p) = E(p_t - p)^2 = p(1-p)\left[1 - \left(1 - \frac{1}{2N_e}\right)^t\right]. \tag{S5}$$

We consider the relevant terms in eq. (S3) separately:

$$
\begin{aligned}
Var(\hat{x}) &= E(Var(\hat{x}|x)) + Var(E(\hat{x}|x)) \\
&= E\left(\frac{x(1-x)}{R_0}\right) + Var(x|p) \\
&= \frac{1}{R_0}[p - E((x - p + p)^2)] + \frac{p(1-p)}{2S_0} \\
&= \frac{1}{R_0}[p - p^2 - E(x - p)^2] + \frac{p(1-p)}{2S_0} \\
&= \frac{1}{R_0}[p(1-p) - Var(x)] + \frac{p(1-p)}{2S_0} \\
&= \frac{1}{R_0}\left[p(1-p) - \frac{p(1-p)}{2S_0}\right] + \frac{p(1-p)}{2S_0} \\
&= p(1-p)\left[\frac{1}{2S_0} + \frac{1}{R_0} - \frac{1}{2S_0 R_0}\right].
\end{aligned}
$$

Note that we assume a diploid population with sample size of $2S_0$ and effective population size of $2N_e$. Replacing $2S_0$ with S_0 leads to the variance of $\hat{x}$ in a haploid population with effective size N_e. We now derive the variance of the sample allele frequency at $T = t$.

$$
\begin{aligned}
Var(\hat{y}) &= E(Var(\hat{y}|y)) + Var(E(\hat{y}|y)) \\
&= E\left(\frac{y(1-y)}{R_t}\right) + Var(y|p) \\
&= \frac{1}{R_t}(p - E(y - p + p)^2) + Var(y|p) \\
&= \frac{1}{R_t}(p - p^2 - Var(y|p)) + Var(y|p) \\
&= \frac{1}{R_t}\left[p(1-p) - p(1-p)\left[1 - \left(1 - \frac{1}{2N_e}\right)^t\left(1 - \frac{1}{2S_t}\right)\right]\right] + p(1-p)\left[1 - \left(1 - \frac{1}{2N_e}\right)^t\left(1 - \frac{1}{2S_t}\right)\right] \\
&= p(1-p)\left[\frac{1}{R_t} - \frac{1}{R_t}\left[1 - \left(1 - \frac{1}{2N_e}\right)^t\left(1 - \frac{1}{2S_t}\right)\right] + \left[1 - \left(1 - \frac{1}{2N_e}\right)^t\left(1 - \frac{1}{2S_t}\right)\right]\right] \\
&= p(1-p)\left[1 - \left(1 - \frac{1}{R_t}\right)\left(1 - \frac{1}{2S_t}\right)\left(1 - \frac{1}{2N_e}\right)^t\right]
\end{aligned}
$$

Here, $2S_j$ is the number of sampled chromosomes in a diploid population at time t. For a haploid population, the factor 2 can be ignored for both the sample size and the effective population size.

We introduce a correction term $C_j := \frac{1}{2S_j} + \frac{1}{R_j} - \frac{1}{2S_j R_j}$ for F_c at generations $j \in \{0, t\}$. For haploid populations an analogous correction term is $C_{j(hapl.)} := \frac{1}{S_j} + \frac{1}{R_j} - \frac{1}{S_j R_j}$ ($j \in \{0, t\}$). Based on the results above, the variance in sample allele frequency at $T = 0$ and t can be written as

$$Var(\hat{x}) = p(1-p)C_0 \, , \tag{S6}$$

and

$$Var(\hat{y}) = p(1-p)\left[1 - (1 - C_t)\left(1 - \frac{1}{2N_e}\right)^t\right] \, . \tag{S7}$$

In the haploid case, we have $Var(\hat{y}) = p(1-p)\left[1 - (1 - C_{t(hapl.)})\left(1 - \frac{1}{N_e}\right)^t\right]$.

Pool-seq data can be over-dispersed, e.g. if different individuals in the pool have different probabilities of contributing the sequenced reads.

If the level of overdispersion can be inferred from the data, then it is possible to account for the additional variance by introducing an overdispersion parameter γ. This parameter can be interpreted as the factor by which the actual variance exceeds the theoretical binomial variance. With γ, our correction term becomes

$$C_{j(disp.)} = \frac{1}{2S_j} + \frac{\gamma}{R_j} - \frac{\gamma}{2S_j R_j}. \tag{S8}$$

Following Waples (1989) the denominator in equation (S3) reduces to

$$E(\hat{z} - \hat{x}\hat{y}) = p(1-p) - Cov(\hat{x}, \hat{y}) \, . \tag{S9}$$

We now investigate the covariance term separately for both sampling plans.

1.2   Sampling plan II

Under sampling plan II, a sample of size S_0 is taken before reproduction and is not replaced in the population. Consequenty, the sample of S_0 individuals and the $2N_e$ chromosomes contributing to the next generation are mutually exclusive, implying that the sample allele frequency at generation 0 is uncorrelated to the sample allele frequency at t, i.e. $Cov(\hat{x}, \hat{y}) = 0$ for sampling plan II. We now plug equations (S6), (S7), and (S9) into equation (S3). With equation (S7), we use the equation (2) from the main text.

$$E(F_c) \approx \frac{Var(\hat{x}) + Var(\hat{y}) - 2Cov(\hat{x}, \hat{y})}{p(1-p) - Cov(\hat{x}, \hat{y})}$$
$$= \frac{Var(\hat{x}) + Var(\hat{y})}{p(1-p)}$$
$$= \frac{p(1-p)C_0 + p(1-p)\left[1 - (1 - C_t)e^{-t/2N_e}\right]}{p(1-p)}$$
$$= C_0 + 1 - e^{-t/2N_e} + C_t e^{-t/2N_e}$$
$$= C_0 + 1 - e^{-t/2N_e} - C_t(1 - e^{-t/2N_e}) + C_t$$
$$= C_0 + C_t + (1 - C_t)(1 - e^{-t/2N_e})$$

By solving this equation and replacing $E(F_c)$ by F_c, under plan II we obtain a measure of variance corrected for the two-step sampling

$$F_c' = \frac{F_c - C_0 - C_t}{1 - C_t} \ . \tag{S10}$$

The effective population size can then be estimated using the method of moments, i.e. equating $F_c' = E[F_c']$ and solving for N_e.

1.3 Sampling plan I

Under sampling plan I, a sample of S_0 individuals is taken after reproduction, so the sample will contain chromosomes that might contribute to the next generation. This implies that the sample allele frequency at generation 0 is positively correlated with the sample allele frequency at t, i.e. $Cov(\hat{x}, \hat{y}) > 0$, because both are derived from the same initial sample at generation 0. Note that the covariance in equation (S3) would be zero, if the allele frequency p_0 of this initial sample were known because sampling itself at the two time points are independent. Similar to Waples (1989), it can be shown that the covariance of $\hat{x}$ and $\hat{y}$ is simply the variance of p_0. Indeed

$$\hat{x} = p + (p_0 - p) + (x - p_0) + (\hat{x} - x) = p_0 + (\hat{x} - p_0) \ , \tag{S11}$$

and

$$\hat{y} = p + (p_0 - p) + (p_t - p_0) + (y - p_t) + (\hat{y} - y) = p_0 + (\hat{y} - p_0). \tag{S12}$$

Thus

$$Cov(\hat{x}, \hat{y}) = Var(p_0) = E(p_0 - p)^2 = \frac{p(1-p)}{2N}, \tag{S13}$$

with N denoting the number of diploid individuals in the initial population, and p is the true allele frequency in the gamete pool preceding generation 0. This leads to the following estimate of $E(F_c)$ for plan I

$$E(F_c) \approx \frac{Var(\hat{x}) + Var(\hat{y}) - 2Cov(\hat{x}, \hat{y})}{p(1-p) - Cov(\hat{x}, \hat{y})}$$

$$= \frac{p(1-p)C_0 + p(1-p)\left[1 - (1-C_t)e^{-t/2N_e}\right] - 2\frac{p(1-p)}{2N}}{p(1-p) - \frac{p(1-p)}{2N}}$$

$$= \frac{C_0 + 1 - e^{-t/2N_e} + C_t e^{-t/2N_e} - \frac{1}{N}}{1 - \frac{1}{2N}}$$

$$= \frac{C_0 + (1 - e^{-t/2N_e}) - C_t(1 - e^{-t/2N_e}) + C_t - \frac{1}{N}}{1 - \frac{1}{2N}}$$

$$= \frac{C_0 + C_t + (1 - e^{-t/2N_e})(1 - C_t) - \frac{1}{N}}{1 - \frac{1}{2N}}$$

Replacing $E(F_c)$ by F_c leads to the corrected estimator of F under plan I

$$F_c' = \frac{F_c \left(1 - \frac{1}{2N}\right) - C_0 - C_t + \frac{1}{N}}{1 - C_t} \ . \tag{S14}$$

As under plan II, a method of moments estimator can now be computed using F_c'. Note that for haploid populations, the covariance in eq. (S13) reduces to $Cov(\hat{x}, \hat{y}) = \frac{p(1-p)}{N}$, and eq. (S14) for haploids will be $F_c' = \frac{F_c\left(1 - \frac{1}{N}\right) - C_{0(hapl.)} - C_{t(hapl.)} + \frac{2}{N}}{1 - C_{t(hapl.)}}$.

References

H. Bastide, A. Betancourt, V. Nolte, R. Tobler, P. Stöbe, A. Futschik, and C. Schlötterer. A genome-wide, fine-scale map of natural pigmentation variation in *Drosophila melanogaster*. *PLoS Genet*, 9(6):e1003534, Jun 2013. doi: 10.1371/journal.pgen.1003534. URL http://dx.doi.org/10.1371/journal.pgen.1003534.

A. Futschik, T. Hotz, A. Munk, and H. Sieling. Multiscale dna partitioning: statistical evidence for segments. *Bioinformatics*, 30(16):2255–2262, Aug 2014. doi: 10.1093/bioinformatics/btu180. URL http://dx.doi.org/10.1093/bioinformatics/btu180.

P. E. Jorde and N. Ryman. Unbiased estimator for genetic drift and effective population size. *Genetics*, 177(2):927–935, Oct 2007. doi: 10.1534/genetics.107.075481. URL http://dx.doi.org/10.1534/genetics.107.075481.

R. Kofler and C. Schlötterer. A guide for the design of evolve and resequencing studies. *Mol Biol Evol*, 31(2):474–483, Feb 2014. doi: 10.1093/molbev/mst221. URL http://dx.doi.org/10.1093/molbev/mst221.

M. Nei and F. Tajima. Genetic drift and estimation of effective population size. *Genetics*, 98(3):625–640, Jul 1981.
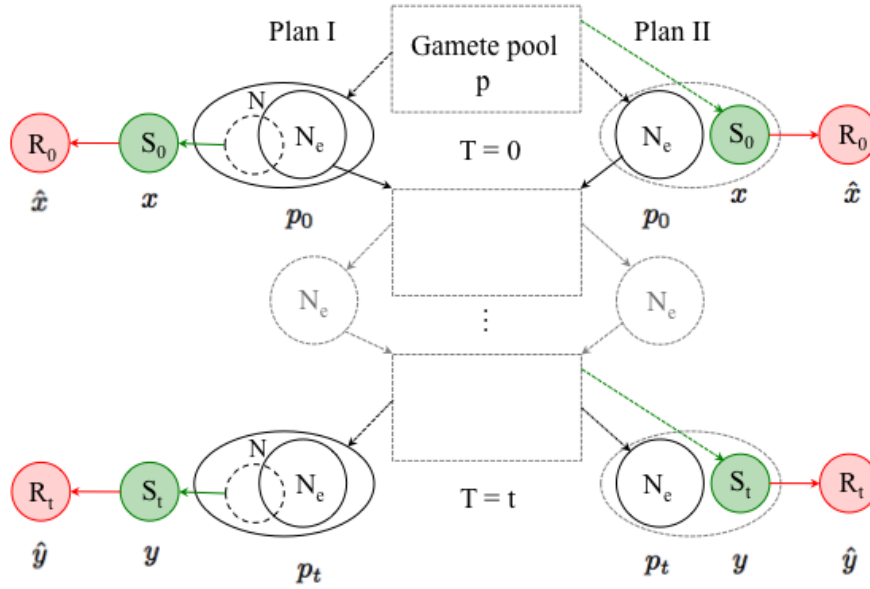
Figure S1: Two-step sampling schemes. We follow the sampling schemes proposed by Nei and Tajima (1981) and generalized by Waples (1989). Under sampling scheme I (plan I) a sample is taken after reproduction. Under the second scheme (plan II), the sample is taken before reproduction and is not returned to the population, so the sampled individuals and the ones contributing to the next generation are mutually exclusive (Waples, 1989). Allele frequencies are obtained after a two-step sampling process: First a sample of S_j ($j \in \{0,t\}$) individuals is taken to create a pooled library for DNA extraction (green), that is subjected to high throughput sequencing. Sampling R_j ($j \in \{0,t\}$) reads is modeled in the second step (red). At each step the relative allele frequencies in the corresponding samples are indicated under the circles representing the sample. This is a modified version of a figure in Waples (1989).
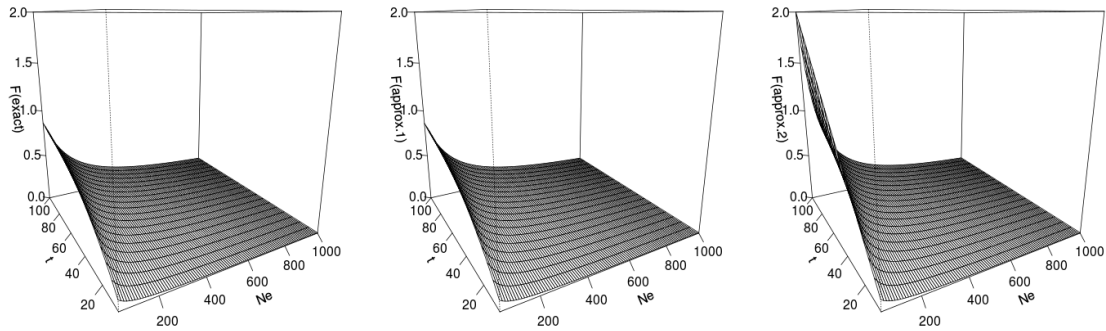
Figure S2: Possible approximations for obtaining our method of moments estimator. The exact method of moments estimator is obtained by solving $F_c' = F := 1 - (1 - 1/2N_e)^t$ for N_e. Left panel: Function plot of F (see main text eq. (1)). Middle: F(approx.1)$= 1 - exp\{-t/2N_e\}$ (main text eq. (2)). Right: F(approx.2)$= t/2N_e$. The function arguments are t and N_e.

R. S. Waples. A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics*, 121(2):379–391, Feb 1989.
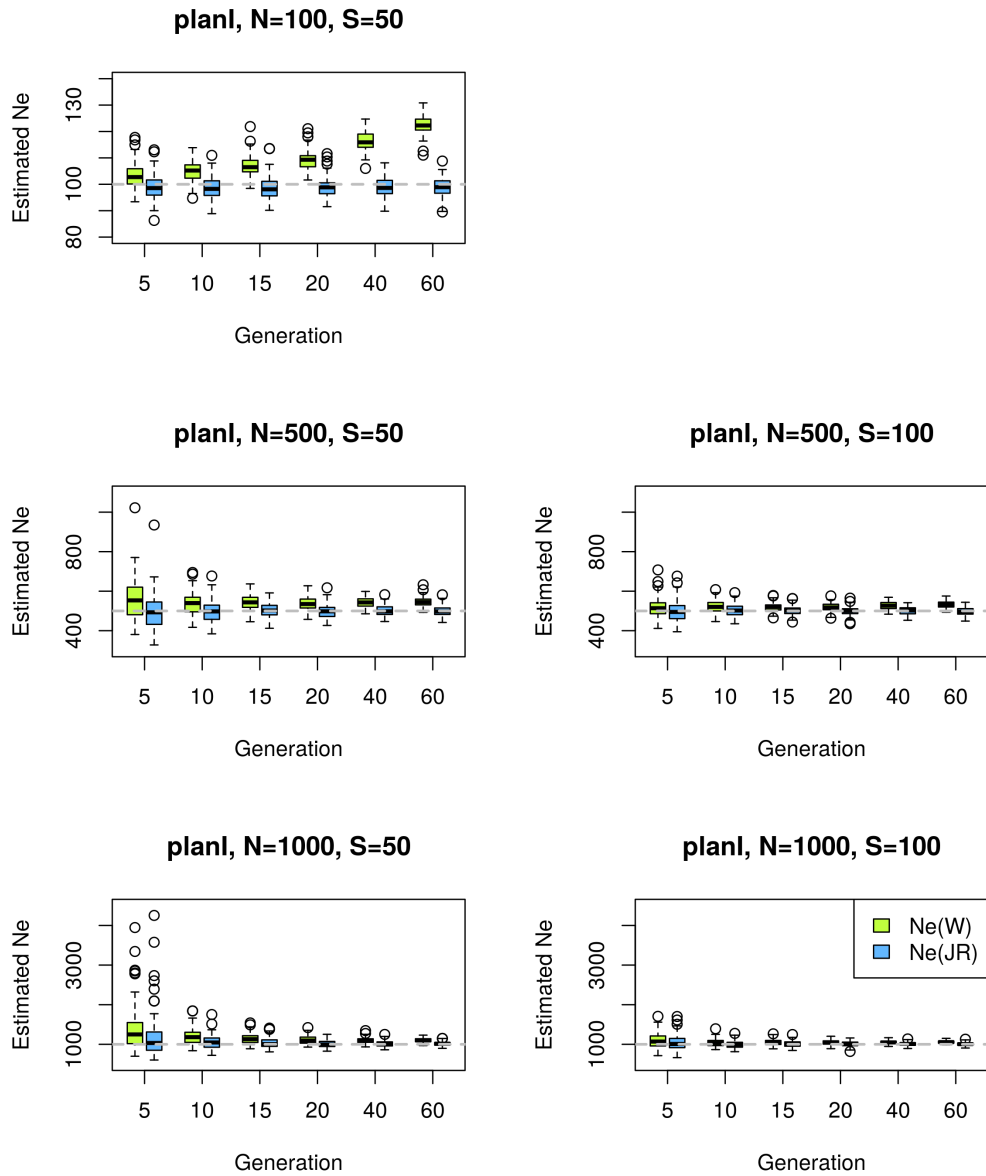
Figure S3: The performance of the estimators proposed by Waples (1989) ($N_e(W)$) and Jorde and Ryman (2007) ($N_e(JR)$) are shown on simulated data. Simulations are performed as described in the main text (section "Simulations"), but without the additional sampling step representing pool sequencing, i.e. full genomic information is known for the S sampled individuals. The true effective population size is indicated by the grey dashed lines We assume $N = N_e$, where N is the census size of the population. For the bottom left scenario the figure is truncated and one outlier estimate > 7000 for $N_e(W)$ is not shown. Only plan I estimators are displayed.
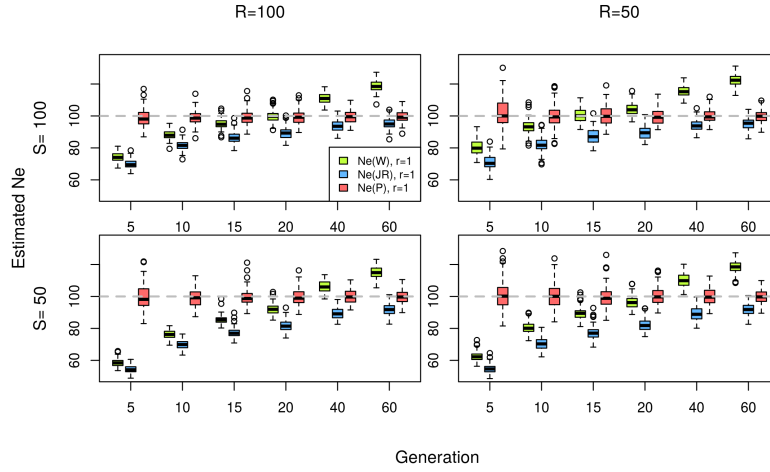
(a) Plan I estimators.
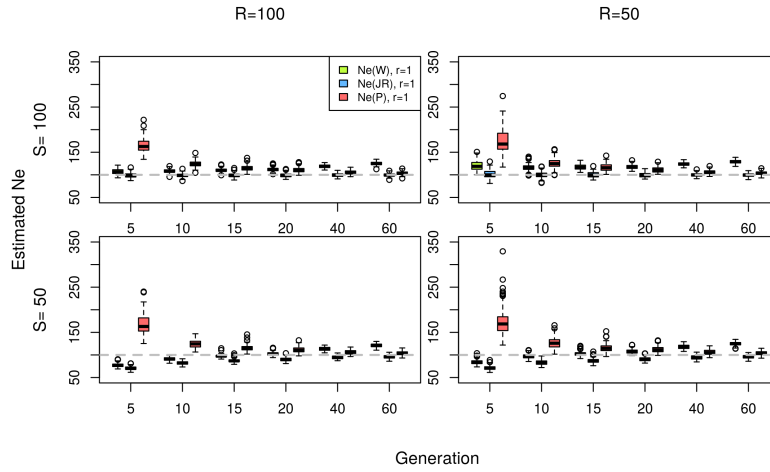
(b) Plan II estimators.

Figure S4: Performance of N_e estimators for $r = 1$. Sixty generations of Wright-Fisher neutral evolution with $N_e = 100$ diploid individuals are simulated for $n = 2000$ loci (SNPs). At the starting population (generation 0), and at each indicated time point, a sample is taken to create a pool of S individuals. Sequencing is modeled by an additional subsequent sampling step of a random number of reads (Poisson with expected value R) for each SNP. Then N_e is estimated on the resulting data set by separately contrasting generation 0 with each of the generations 5, 10, 15, 20, 40 and 60. We consider the estimators $N_e(W)$, $N_e(JR)$ and $N_e(P)$. Each box represents results from 100 simulations with identical parameters. The grey dashed lines indicate the actual value of N_e. Results under plan I and II are shown separately on panels (a) and (b), respectively.

9
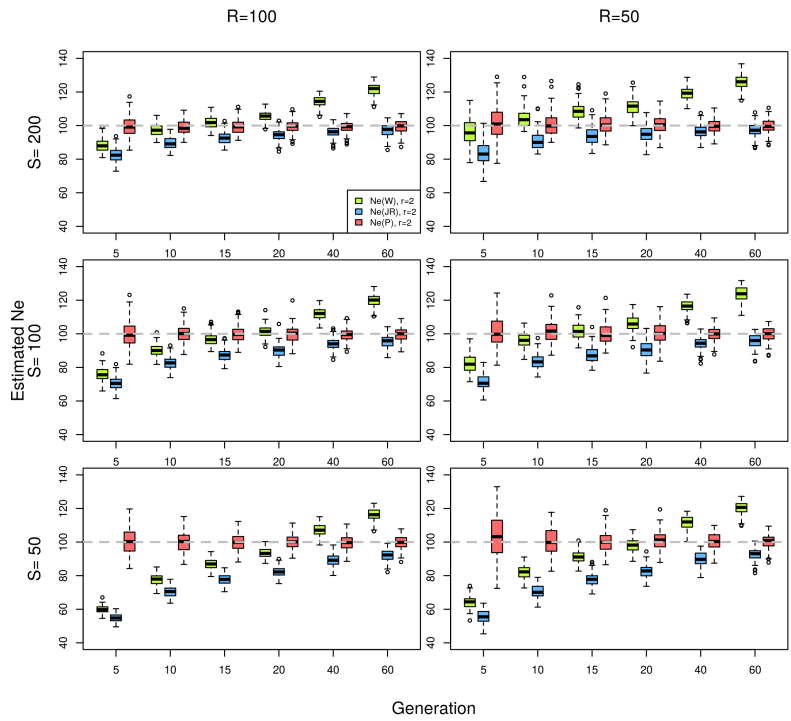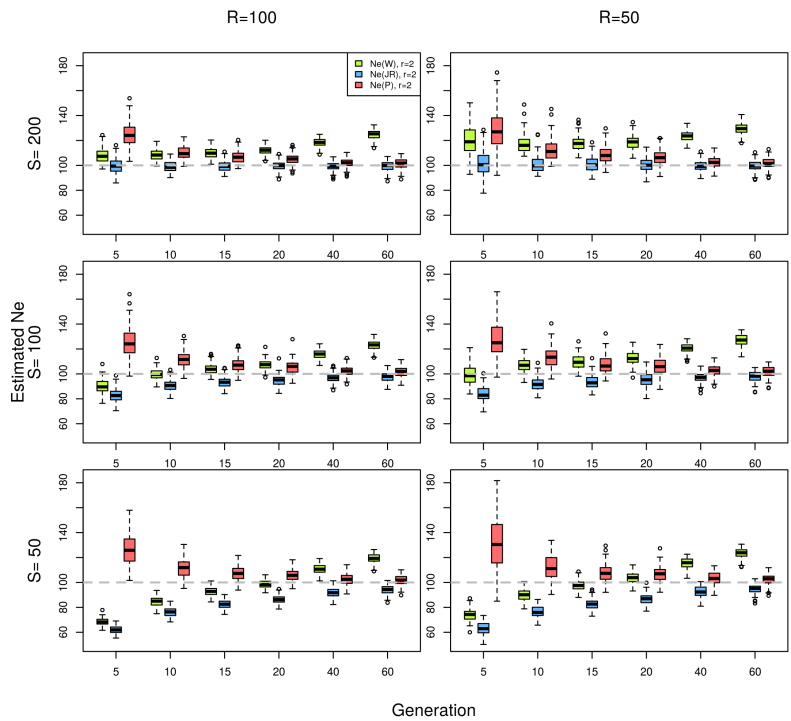
(a) Plan I estimators

(b) Plan II estimators

10

Figure S5: Performance of N_e estimators for $r = 2$. Data simulated as described in Fig. S4.
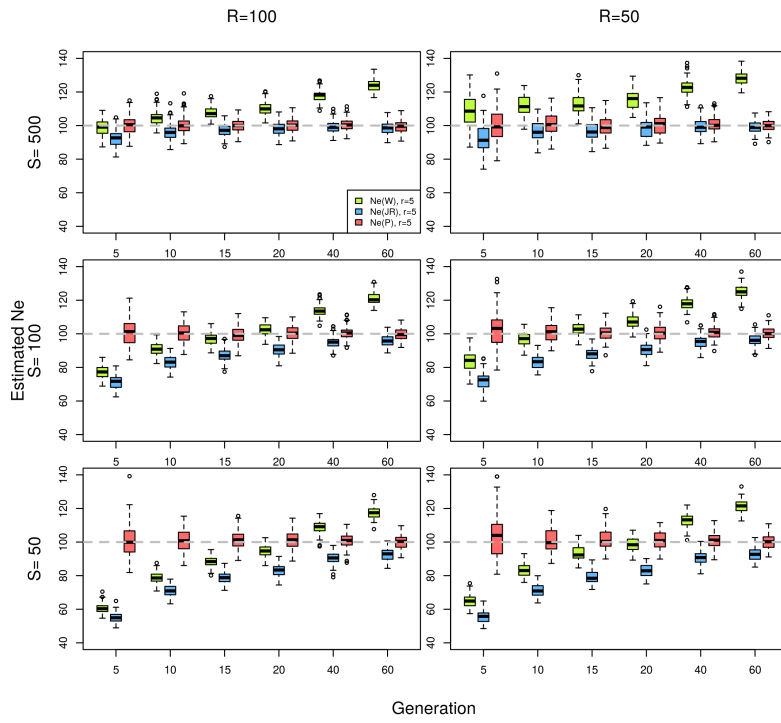
(a) Plan I estimators
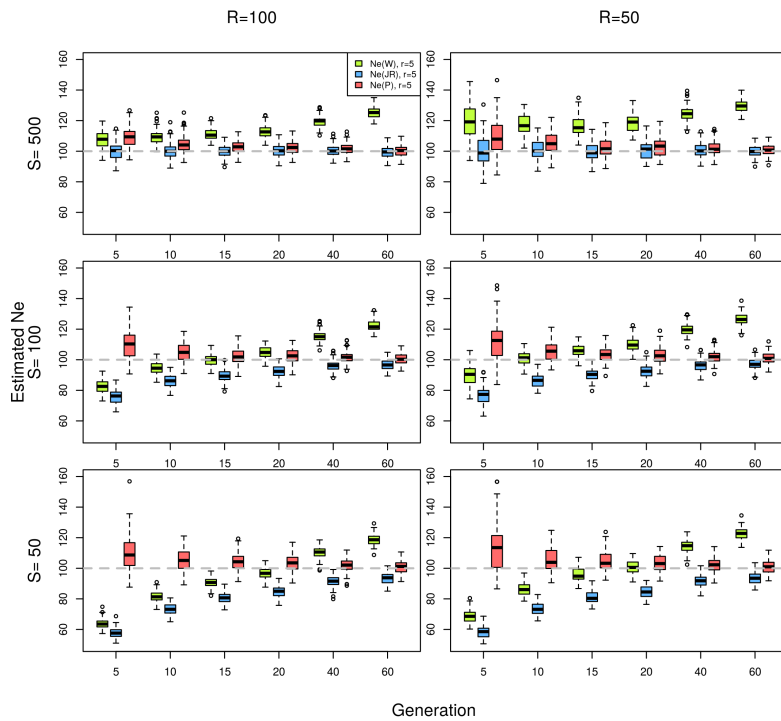
(b) Plan II estimators

Figure S6: Performance of N_e estimators for $r = 5$. Data simulated as described in Fig. S4.
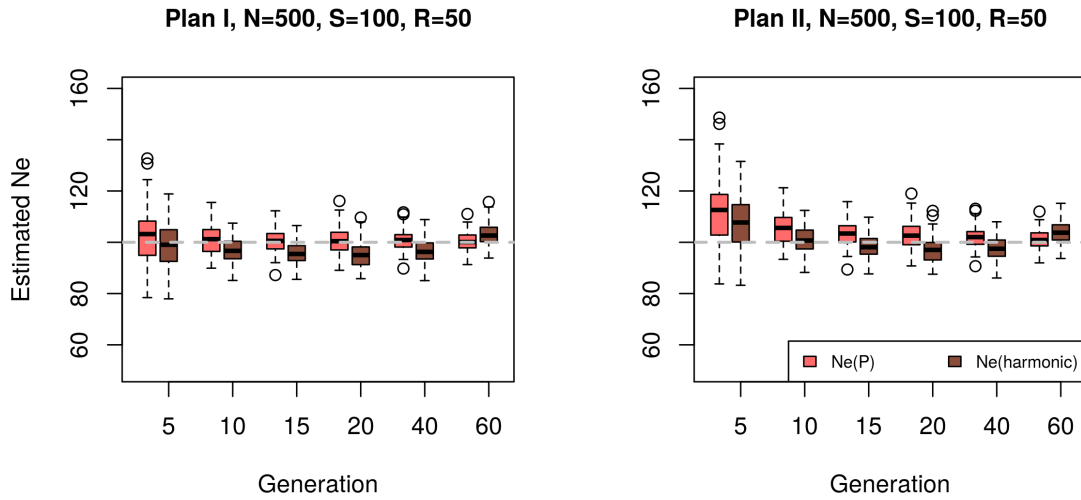
Figure S7: An alternative estimate of N_e ($N_e^*(P)$), that is defined as the harmonic mean of individual N_e estimates (one per locus). Before taking the harmonic mean, our two-step correction is applied to all individual estimates. $N_e(P)$ is the original estimator also shown in Fig. 2. Simulation parameters are matching that of Fig. 2.

	Median					
Repl.	X	2L	2R	3L	3R	Genome-wide
R1	254.7478	235.0195	264.2104	185.1178	125.1865	191.5493
R2	356.6835	299.7259	277.5293	195.7418	181.2761	244.5242
R3	257.3661	255.4859	208.1953	149.7616	131.1328	160.2565

Table S1: Genome-wide median $\hat{N}_e$ from an E&R study with *D. melanogaster*. The effective population size is estimated with $N_e(P)$ plan I in windows of 10000 SNPs (Fig. 7). The median across windows of the estimates is shown for the major chromosome arms. The genome-wide median is taken over the autosomes excluding chromosome 4.
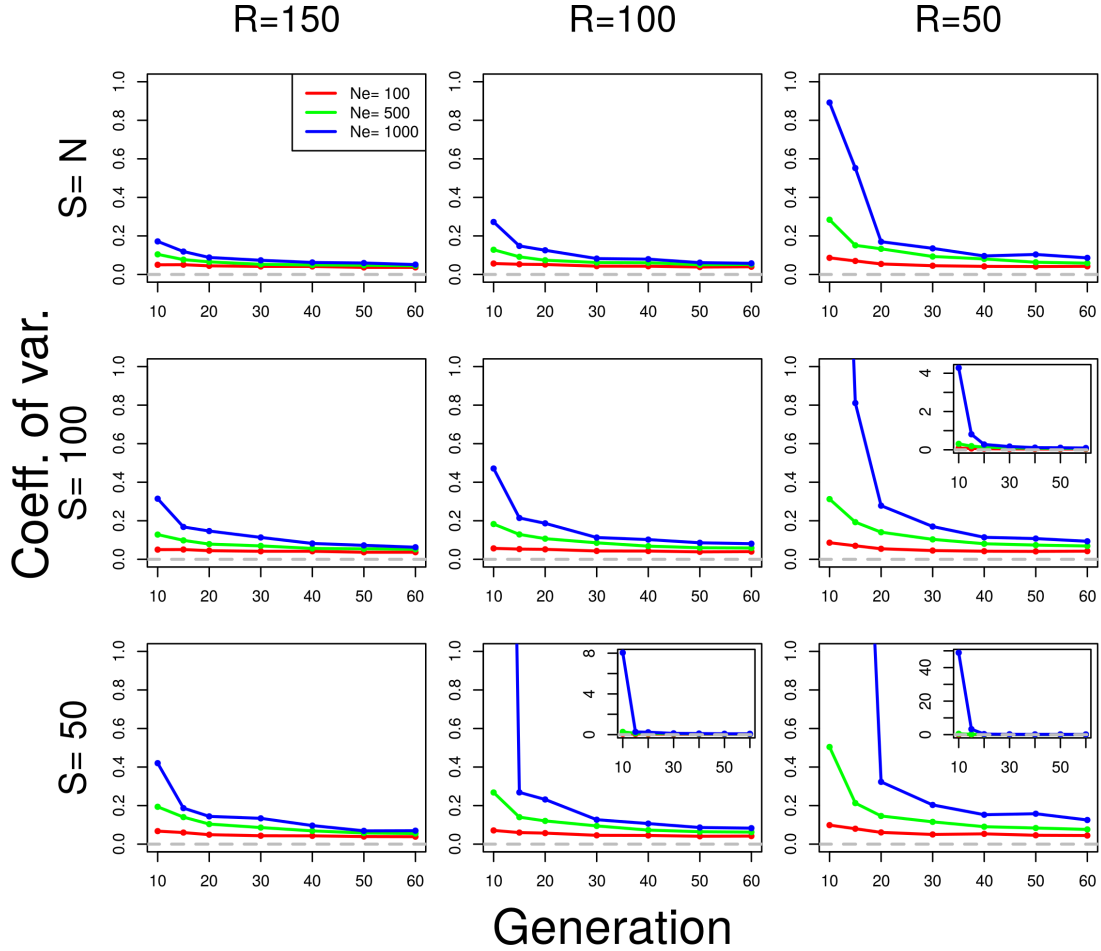
Figure S8: Coefficient of variation of $N_e(P)$ under plan II for various parameter values. Neutral Wright-Fisher simulations were performed with various combinations of the parameters: effective population size ($N_e = 100, 500, 1000$ diploid individuals), pool size ($S = 100, 50$) and coverage ($R = 150, 100, 50$). N_e is estimated with $N_e(P)$ under plan II using $n = 2000$ SNPs. $S = N$ indicates scenarios when the whole population is sequenced as a single pool. For all simulations, $N = N_e$ is used. Each value is calculated over 100 simulations. When the coefficient of variation exceeds one, an inset figure shows the actual value.
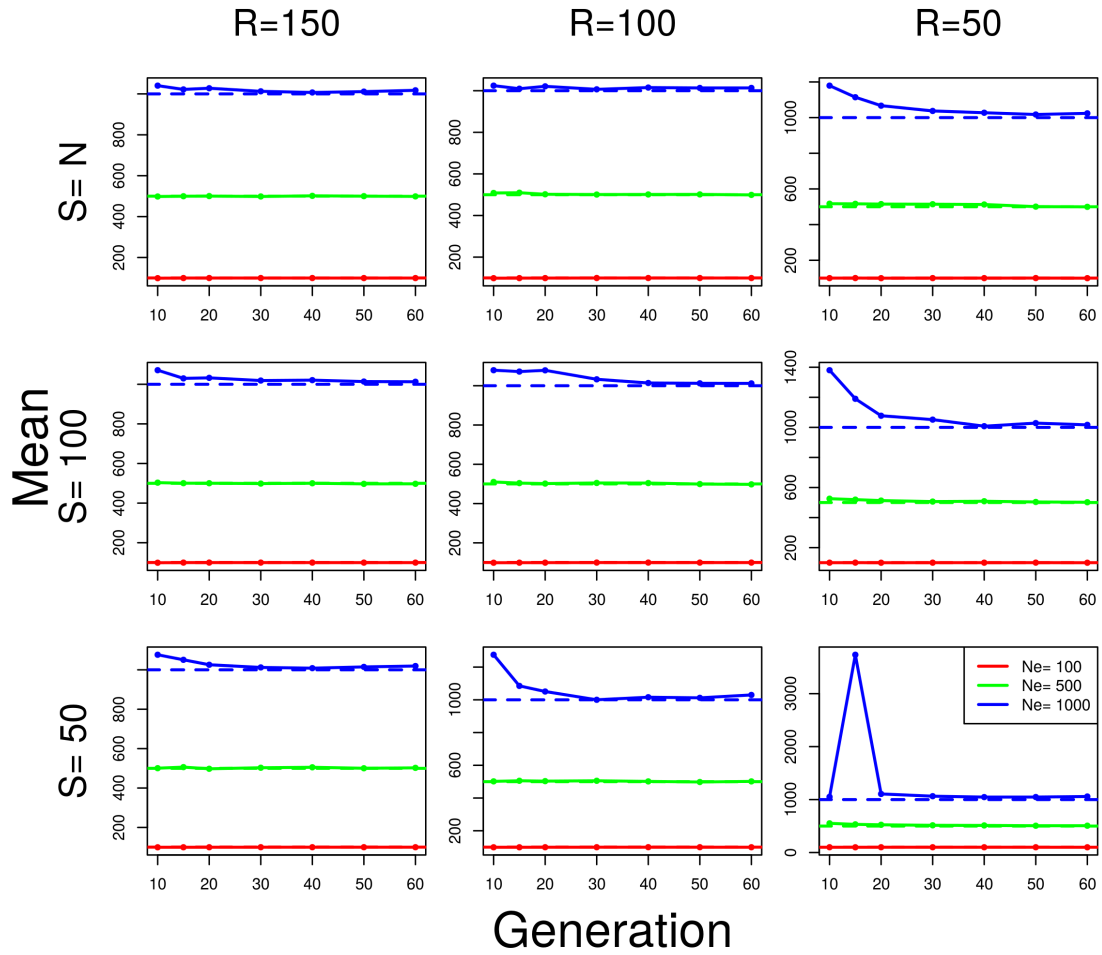
Figure S9: Mean of $\hat{N}_e$ estimated with $N_e(P)$ under plan I for various parameter values. Wright-Fisher neutral simulations were performed as for Fig. S8. The dashed lines indicate the actual value of N_e. Panels can have different y-axis scale.
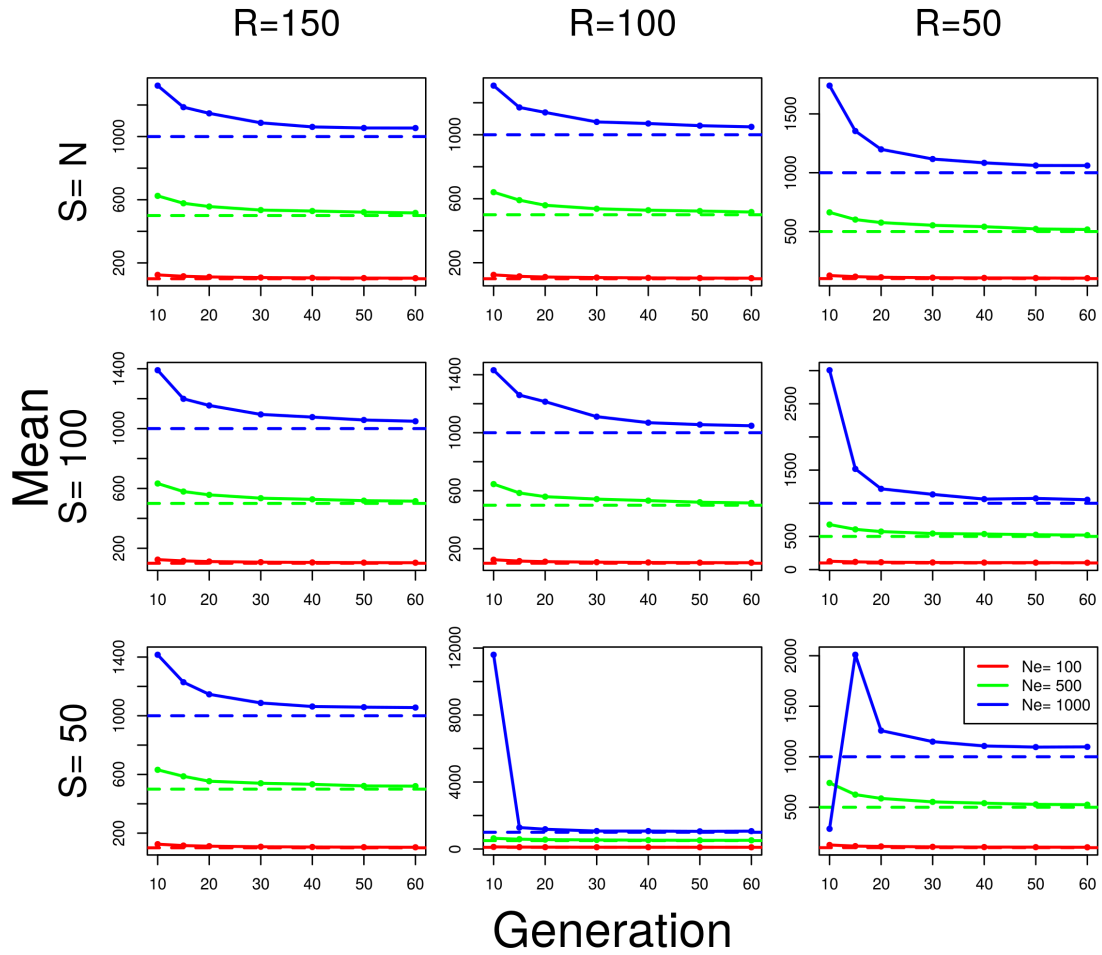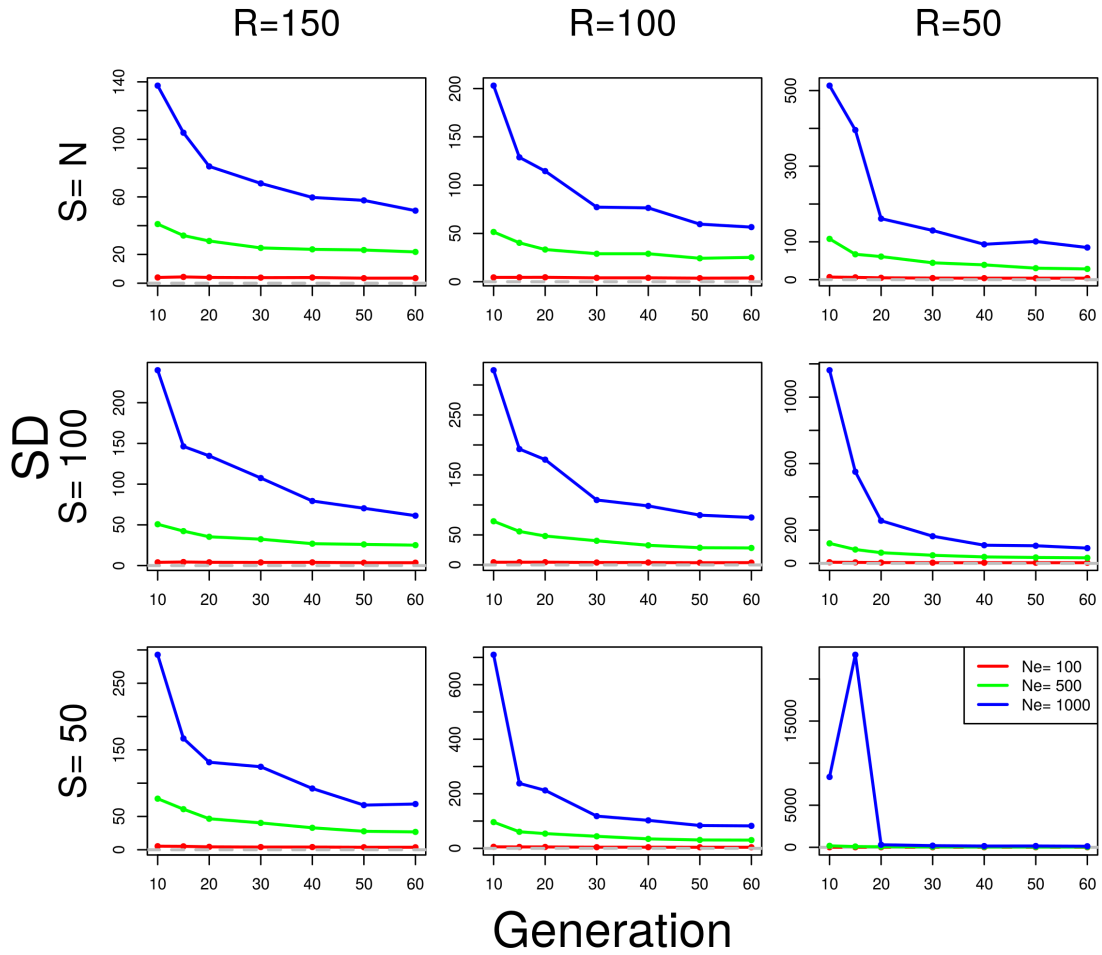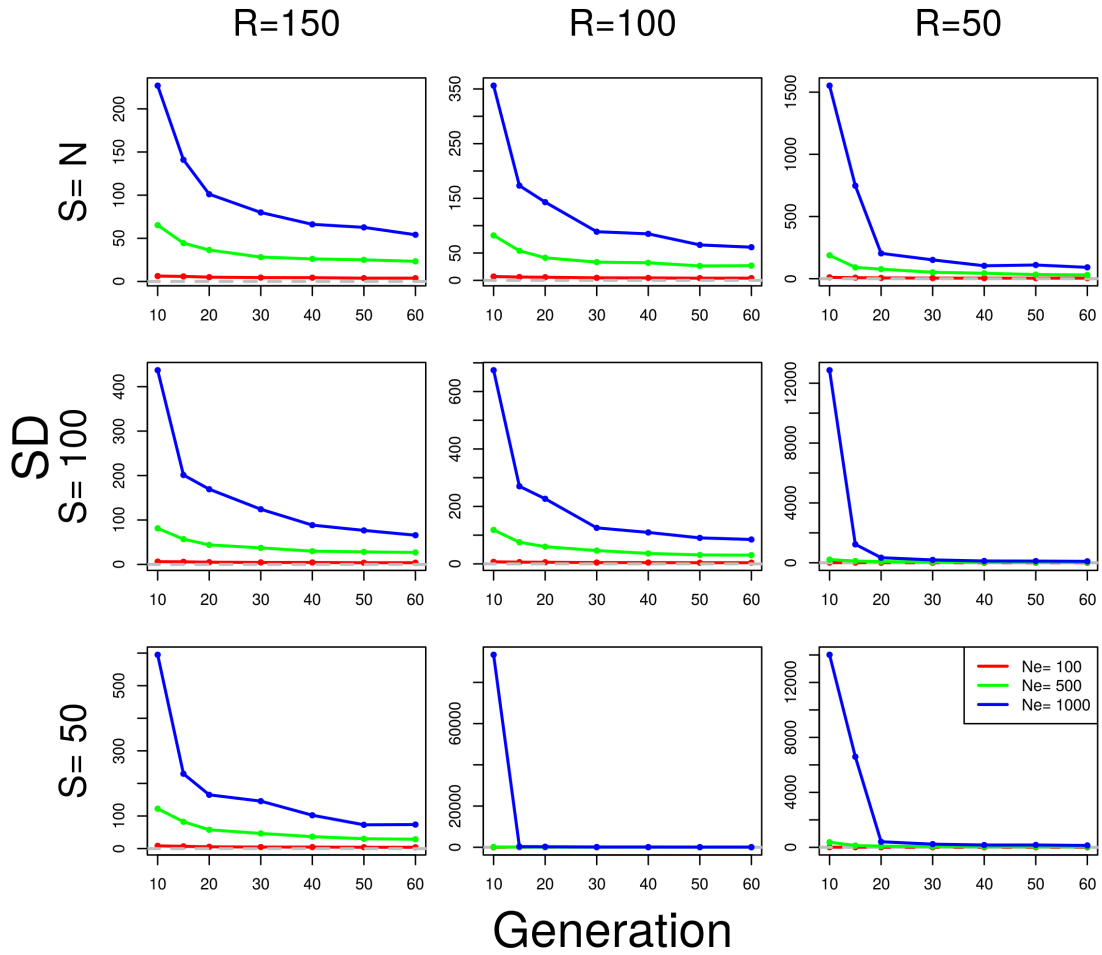
Figure S10: Mean of $\hat{N}_e$ estimated with $N_e(P)$ under plan II for various parameter values. Wright-Fisher neutral simulations were performed as for Fig. S8. The dashed lines indicate the actual value of N_e. Panels can have different y-axis scales.

Figure S11: Standard deviation of $\hat{N}_e$ estimated with $N_e(P)$ under plan I for various parameter values. Wright-Fisher neutral simulations were performed as for Fig. S8. Panels can have different y-axis scales.
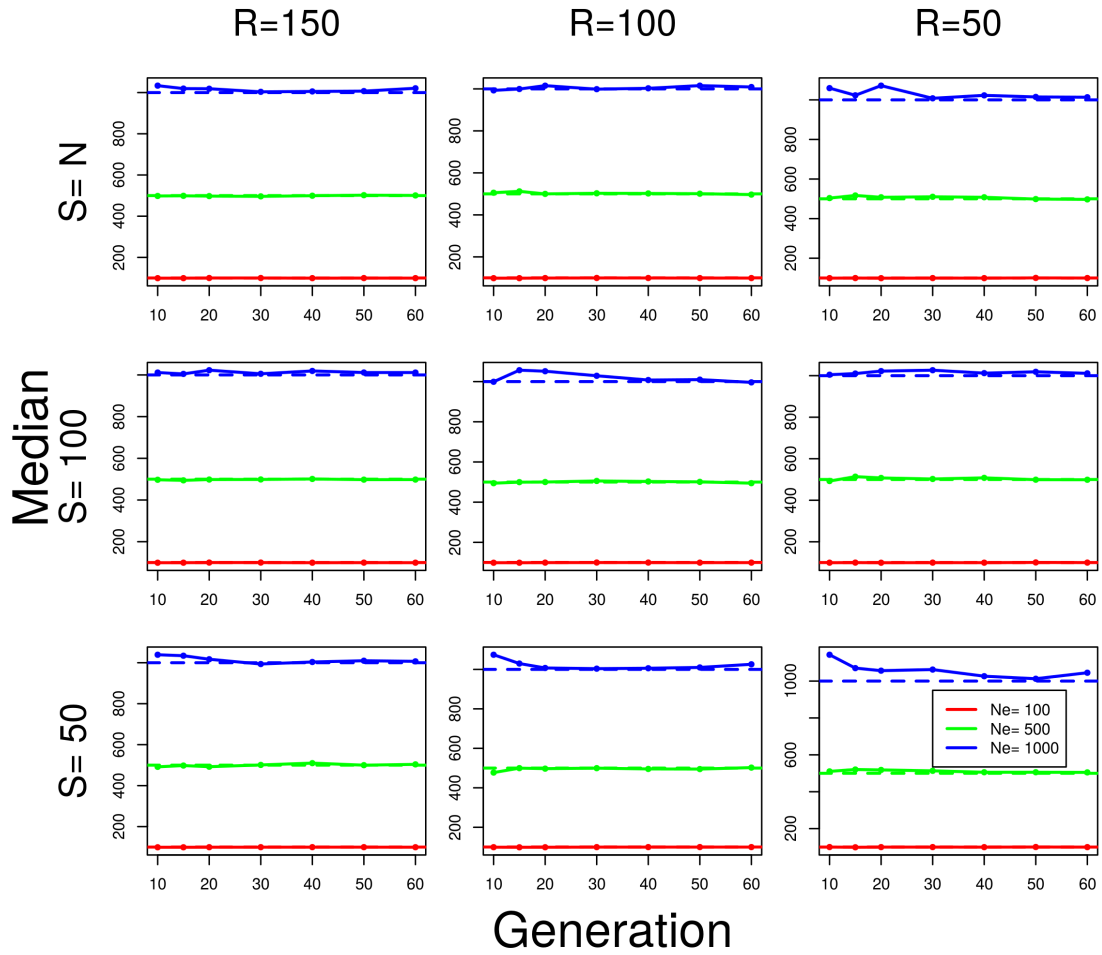
Figure S12: Standard deviation of $\hat{N}_e$ estimated with $N_e(P)$ under plan II for various parameter values. Wright-Fisher neutral simulations were performed as for Fig. S8. Panels can have different y-axis scales.

Figure S13: Median of $\hat{N}_e$ estimated with $N_e(P)$ under plan I for various parameter values. Wright-Fisher neutral simulations were performed as for Fig. S8. The dashed lines indicate the actual value of N_e.
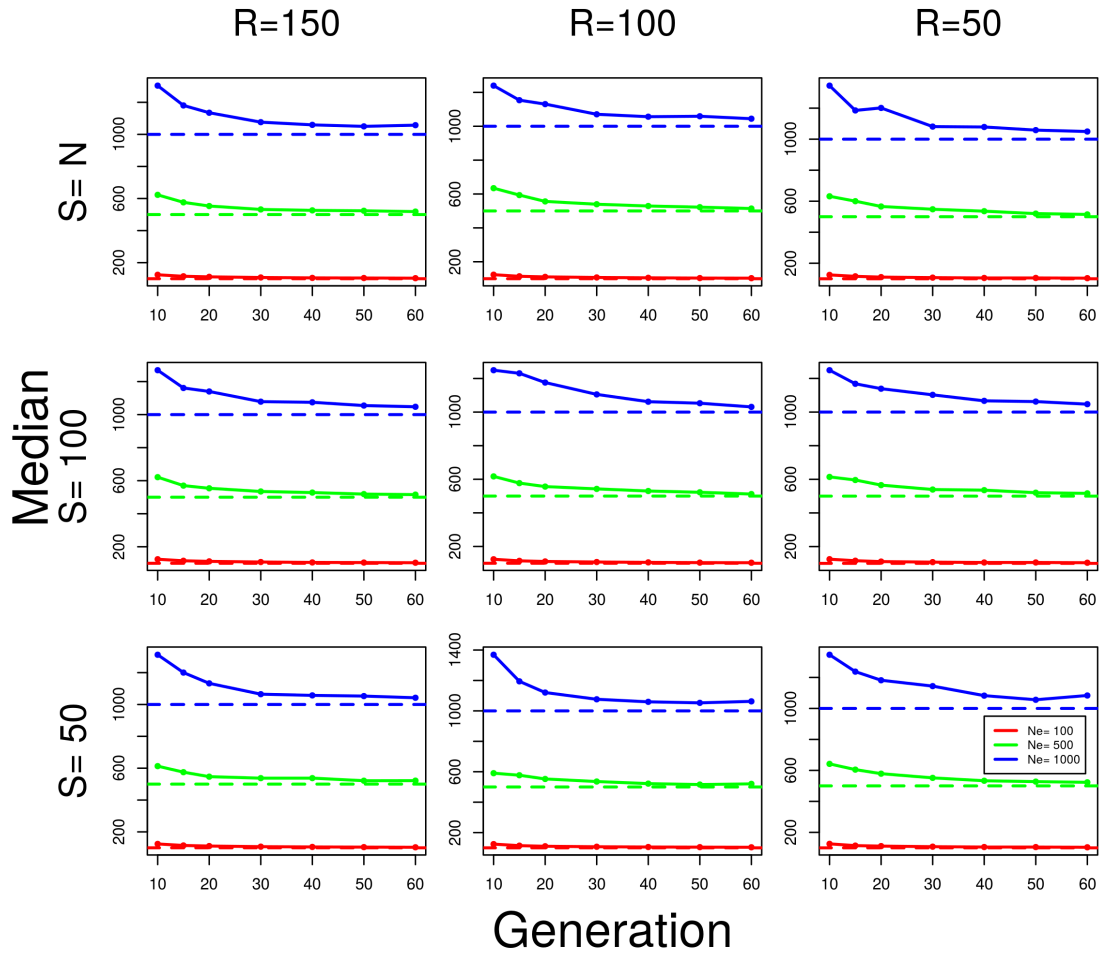
Figure S14: Median of $\hat{N}_e$ estimated with $N_e(P)$ under plan II for various parameter values. Wright-Fisher neutral simulations were performed as for Fig. S8. The dashed lines indicate the actual value of N_e.
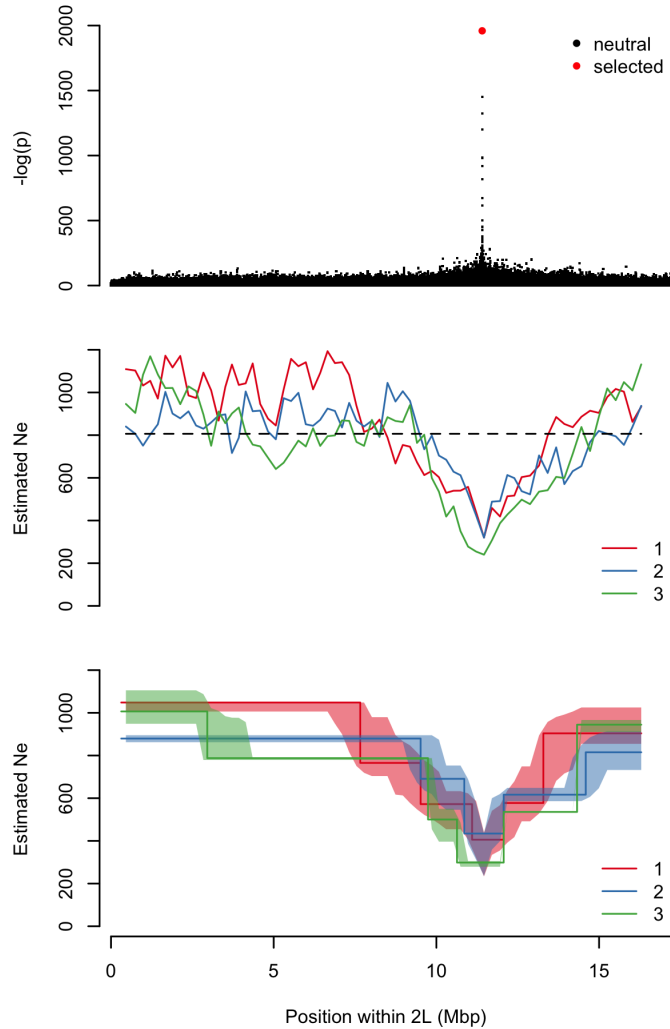
Figure S15: Estimated N_e on data simulated with selection. We performed forward Wright-Fisher simulations for 59 generations with selection in three replicate populations using MimicrEE (Kofler and Schlötterer (2014). (See also Section Simulations in the Main text). A selection coefficient of $s = 0.2$ (with semi-dominance, $h = 0.5$) is assigned to a single randomly selected locus with low starting allele frequency. All other loci were evolving neutrally. The population size is kept at a constant level of 1000 individuals during the simulations. MimicrEE simulates haplotype evolution with recombination, reflecting linkage disequilibrium in the data. The top panel shows the p-values of a selection scan indicating the true target of selection in red (Cochran-Mantel-Haenszel test contrasting allele frequencies at generation 0 and 59, Bastide et al. (2013)). N_e is estimated in non-overlapping windows containing 10000 SNPs independently for replicates (middle panel). The mean $\hat{N}_e$ across windows and replicates is shown with the dashed line. The results of the segmentation algorithm performed on $\hat{N}_e$ windows (Futschik et al., 2014) is shown in the bottom panel separately for the replicates.
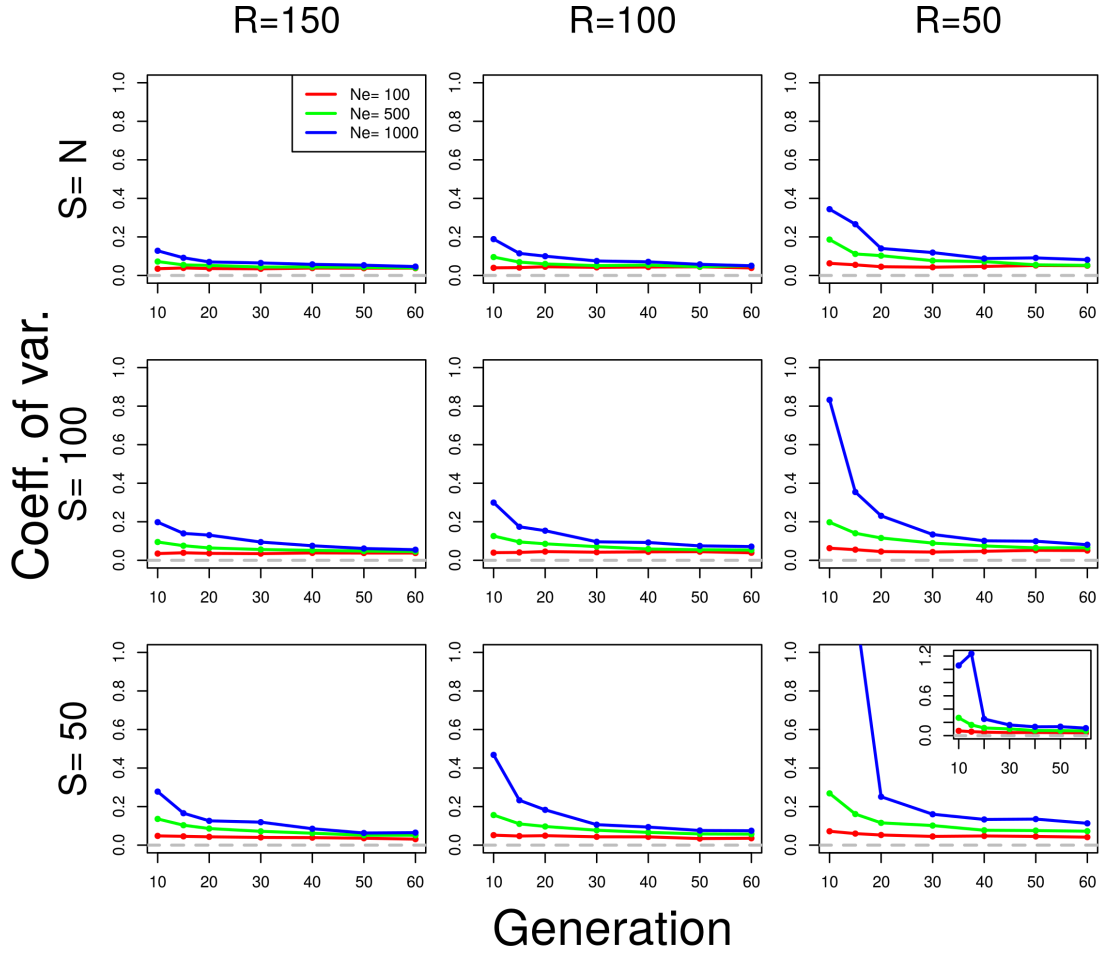
Figure S16: Coefficient of variation of $\hat{N}_e$ estimated with the harmonic mean estimator $N_e^*(P)$ under plan I for various parameter values. Wright-Fisher neutral simulations were performed as in Fig. S8. Effective population size is estimated separately for each marker locus using the corrected variance estimator in eq. (11). $\hat{N}_e$ is then estimated as the harmonic mean across the n single locus N_e estimates.
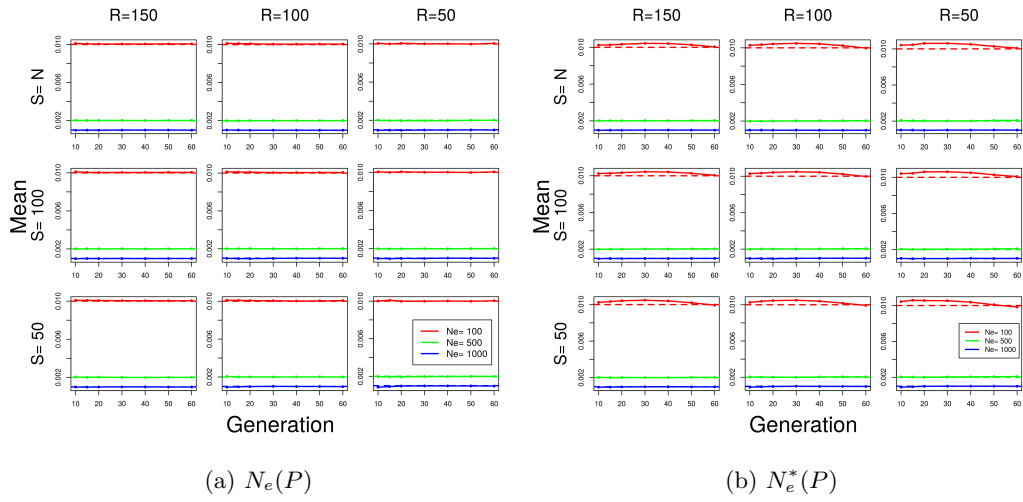
(a) $N_e(P)$
(b) $N_e^*(P)$

Figure S17: Mean of the reciprocal of estimators $N_e(P)$ and $N_e^*(P)$. Wright-Fisher neutral simulations were performed as in Fig. S8. (a) Mean of $1/\hat{N_e}$, where $\hat{N_e}$ is estimated with $N_e(P)$. (b) Mean of $1/\hat{N_e}$, where $\hat{N_e}$ is calculated with $N_e^*(P)$.