

Estimating effective population size from temporal allele frequency changes in experimental evolution

Ágnes Jónás^{*†}, Thomas Taus^{*†}, Carolin Kosiol[†], Christian Schlötterer[†], Andreas Futschik^{‡†}

May 4, 2016

^{*}Vienna Graduate School of Population Genetics, 1210 Wien, Austria

[†]Institut für Populationsgenetik, Vetmeduni Vienna, 1210 Wien, Austria

[‡]Department of Applied Statistics, Johannes Kepler Universität Linz, 4040 Linz, Austria

Running Head: Estimate N_e in experimental evolution

Key Words: Effective population size, Genetic drift, Pool-Seq, Experimental evolution

Corresponding Author:

Andreas Futschik

Department of Applied Statistics

Johannes Kepler Universität Linz

Altenberger Straße 69

4040 Linz, Austria

`andreas.futschik@jku.at`

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 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 increasing the estimated allele frequency variance. 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 type I error is accounted for, our method permits the identification of genomic regions that differ significantly in N_e . 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>.

INTRODUCTION

During experimental evolution studies, populations are maintained for many generations under specific laboratory conditions (KAWECKI *et al.* 2012; SCHLÖTTERER *et al.* 2015; LONG *et al.* 2015). For sexual organisms, such as *Drosophila*, the founder population is derived from a natural population and maintained in a new environment, typically at a census size of less than 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 effective population size (N_e) (WRIGHT 1931). Furthermore, under certain conditions, the efficacy of deterministic forces, such as selection, is controlled by the population size (CHARLESWORTH 2009). For weakly selected alleles the probability of fixation is directly proportional to the product of N_e and intensity of selection (FISHER 1930; KIMURA 1964; KIMURA 1964). Since changes in allele frequency are greatly affected by the population size, it is fundamental to estimate N_e in order to understand molecular variation of experimental evolution studies.

KRIMBAS and TSAKAS (1971) estimated N_e using the standardized variance of allele frequency (F) from longitudinal samples in natural populations of olive flies. The underlying idea is that under neutral Wright-Fisher evolution, the variance of allele frequency after t generations is well described by the starting allele frequency and the drift level, which is inversely proportional to N_e (see also FALCONER and MACKAY (1996)). However, the frequencies are estimated from a sample, such that F must be corrected for the random errors due to sampling from the entire population. NEI and TAJIMA (1981) derived N_e estimates under two different sampling plans. NEI and TAJIMA (1981) and POLLAK (1983) developed improved measures of F with reduced variance. Finally, WAPLES (1989) introduced a generalized framework that unified the different sampling plans introduced by NEI and TAJIMA (1981). Using their framework JORDE and RYMAN (2007) proposed an alternative estimator of N_e that weights alleles differently and reduces the bias of the previous estimators.

With the advent of powerful computers, maximum likelihood-based methods became in-

creasingly popular (WILLIAMSON and SLATKIN 1999; ANDERSON *et al.* 2000; WANG 2001; HUI and BURT 2015) in addition to the moment-based approaches discussed above. Likelihood methods have the advantage that they can be extended beyond examining neutrally evolving isolated populations. For example, they permit to include migration (WANG and WHITLOCK 2003) and selection (BOLLBACK *et al.* 2008; MALASPINAS *et al.* 2012; MATHIESON and MCVEAN 2013). 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; WAPLES and DO 2010; WAPLES and ENGLAND 2011), heterozygote excess (PUDOVKIN *et al.* 1996), molecular co-ancestry (NOMURA 2008), sibship frequencies (WANG 2009; WANG 2013) or combination of summary statistics using approximate Bayesian computation (TALLMON *et al.* 2008). LD-based methods require haplotype data, which limits their applicability to large-scale population studies.

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.*

2011; KOFLER *et al.* 2011; KOLACZKOWSKI *et al.* 2011; FERRETTI *et al.* 2013; BOITARD *et al.* 2013). To the best of our knowledge, no N_e estimators have been used so far which account for the peculiarities of Pool-Seq.

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 biases, as they neglect the variance component due to pooled sequencing. We introduce a new model accounting for the two stage 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 classical temporal N_e estimators. We then extend our estimators to heterogeneous effective population sizes along a chromosome. Partitioning DNA sequences into a number of segments is used to obtain separate N_e estimates for each segment. Finally, we present an application to a genome-wide experimental evolution data set from *Drosophila melanogaster* (FRANSSEN *et al.* 2015).

MATERIALS AND METHODS

Two sampling schemes NEI and TAJIMA (1981) pointed out that our ability to accurately estimate the population size depends on assumptions about the method of sampling individuals for genetic analyses. They proposed two different sampling schemes. Under the first scheme (plan I), individuals are either sampled after reproduction, or returned to the population after examining their genotypes. NEI and TAJIMA (1981) used the hypergeometric distribution to model draws of individuals from a starting population. This restricts the analysis to the assumption that actual size of the population is equal to the effective size. 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. The authors applied plan II assuming that the ratio between census and effective population size is large, so that sampling does not have

an effect on the underlying allele frequency dynamics. In such situations binomial sampling can be used as an approximation.

WAPLES (1989) considered 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 generation 0 and t are uncorrelated.

Contrary to the previous approaches, we consider two distinct sampling steps at each time point (Fig. 1). In the first step, we sample individuals out of the population to create pooled samples for sequencing. Sampling individuals can take place according to either plan I or plan II. In the second step, we model drawing reads out of the pooled DNA sample. The estimated allele frequency variance is then corrected for the additional variance coming from both sampling steps.

For a typical E&R study, outbred experimental populations are created by mixing a large number of inbred lines (TURNER and MILLER 2012; BASTIDE *et al.* 2013; 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 (N) controlled through time (Fig. 1). However, the experimenter has no direct influence on the effective populations 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. Sampled individuals are subsequently pooled together. 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. Sequenced reads are mapped to a reference genome and filtered for subsequent analyses. Randomness in sequencing and local structures in the genome lead to a variable coverage level, which we also take into account when estimating N_e .

[Figure 1 about here.]

Notation We use observed allele frequencies sampled t generations apart to estimate N_e (Fig. 1). We assume that the population is propagated at constant census size N , and that $N \geq N_e$. We consider only biallelic sites. At each locus the true population allele frequency at time $T = j$ is denoted by p_j , where $j \in \{0, t\}$. To obtain allele frequency estimates for the unknown p_j , the population is subjected to sampling. In contrast to previous approaches, we consider two sampling steps (Fig. 1). At $T = j$, we first sample S_j individuals out of the population to create a pooled sample for sequencing. Sampling can take place according to either plan I or plan II, as described above (also shown in Fig. S1). For the second sampling step, we model Pool-Seq by drawing R_{ij} reads out of the pooled DNA sample at site i . This permits for variation in sequence coverage. Below we 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 following expression

$$\sigma_p^2 = (1 - p_0)p_0 \left[1 - \left(1 - \frac{1}{2N_e} \right)^t \right], \quad (1)$$

where p_0 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

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

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

$$F = F_a := \frac{1}{a} \sum_{i=1}^a \frac{(x_i - y_i)^2}{x_i(1-x_i)} , \quad (3)$$

where x_i and y_i ($i = 1, \dots, a$) 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 binomial 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 unbiased estimator to measure temporal frequency change.

Correction for two stage sampling We follow the sampling approach described above, when calculating the correction term for pooled samples. We consider a diploid, random mating population of size N with discrete generations. Neutral evolution is assumed with no selection, migration and mutation. Samples are drawn from the population at generation

$T = 0$ and t . To model sampling variation, we assume two stage sampling for both sampling plans as illustrated in Fig. S1. In the first step, we model drawing S_j individuals out of N to create pooled sequencing samples. We consider diploid populations throughout the derivation, so that a sample of S_j individuals leads to $2S_j$ sequences. Sampling is assumed to be binomial with parameters $2S_j$ and p_j (WAPLES 1989). In the second step, pool sequencing a random number R_j of reads is also modeled as binomial sampling.

By calculating the variance of the observed allele frequencies in pooled samples, we are able to correct for the variation introduced by the two-step sampling. Following JORDE and RYMAN (2007), we use the following expression as the measure for the temporal change in allele frequency

$$F_c = \frac{\sum_{i=1}^a (\hat{x}_i - \hat{y}_i)^2}{\sum_{i=1}^a \hat{z}_i - \hat{x}_i \hat{y}_i}, \quad (4)$$

where $\hat{z}_i = (\hat{x}_i + \hat{y}_i)/2$ and a is the number of alleles (NEI and TAJIMA 1981). We consider only biallelic sites, which reduces F_c to $(\hat{x} - \hat{y})^2/(\hat{z} - \hat{x}\hat{y})$ for a locus. The expectation of F_c is approximated in the following way

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

For both plans, we derive expressions for the numerator and denominator in equation (5) separately. Here we summarize our main conclusions, details on the derivation are provided in the supplementary material. With $C_j := \frac{1}{R_j} + \frac{1}{2S_j} - \frac{1}{R_j 2S_j}$ for $j \in \{0, t\}$, and p denoting the true population allele frequency in the gamete pool at generation 0, we obtain

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

and

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

Following WAPLES (1989) the denominator in equation (5) reduces to

$$E(\hat{z} - \hat{x}\hat{y}) = p(1 - p) - Cov(\hat{x}, \hat{y}) . \quad (8)$$

For plan II, the sample allele frequencies at generation 0 and t are uncorrelated, $Cov(\hat{x}, \hat{y}) = 0$ and F_c corrected for the noise coming from the two stage sampling is given by

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

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 to calculate the sample covariance $Cov(\hat{x}, \hat{y})$ in equation (5). It turns out (see supplementary material for details) that the covariance of \hat{x} and \hat{y} is equal to the variance of p_0

$$Cov(\hat{x}, \hat{y}) = Var(p_0) = E(p_0 - p)^2 = \frac{p(1 - p)}{2N}, \quad (10)$$

where N is the census size of the population at generation 0. Substituting the inferred covariance into equation (5) leads to the following corrected variance estimate, F'_c for plan I

$$F'_c = \frac{F_c \left(1 - \frac{1}{2N}\right) - C_0 - C_t + \frac{1}{N}}{1 - C_t} . \quad (11)$$

N_e is then estimated using the corrected variance estimates (equations (9) and (11)) instead of F in equation (2).

Having long time series, often spanning hundreds of generations, has recently become commonplace (BARRICK *et al.* 2009; BURKE *et al.* 2010; BURKE *et al.* 2014). The assumption of t being small, often made when calculating N_e (KRIMBAS and TSAKAS 1971; NEI and TAJIMA 1981), can lead to severe bias when approximating $2N_e \approx t/F$. Therefore, instead of assuming t being small, we use the formula in equation (2) for estimating N_e .

In our genome-wide data set, we apply individual correction terms C_{ij} ($j \in \{0, t\}$) for each marker i ($i = 1, \dots, n$) before averaging over the n biallelic loci in a window.

Simulations We evaluate the performance of our novel estimator on simulated data. Using the neutral Wright-Fisher model, we simulate allele frequency trajectories at n independent loci (SNPs) in a population of N_e diploid individuals. Usually our simulations start with uniformly distributed allele frequencies. Nevertheless, we also assess the performance of our estimator on simulated data with a starting allele frequency spectrum skewed towards low frequency variants resembling neutral expectations. To investigate the effect of the ratio between census and effective population size ($r = N/N_e$), we increment the population size to a desired level of N while keeping the allele frequencies unchanged before sampling takes place. At the start and after t generations samples are taken out of the population. Pooling individuals is modeled by sampling without replacement. Genome-wide sequence data typically has uneven coverage among sites, which we model by a Poisson distribution with a given mean coverage. For every biallelic genomic position, the number of reads corresponding to the target coverage is determined by binomial sampling. N_e is estimated between the start and after t generations for $n=1000$ SNPs in 100 replicate simulations. We report summary statistics of the resulting N_e estimates across replicates.

Linkage disequilibrium between loci can reduce the number of independent SNPs, thereby increasing the variance of the estimate. The impact of non-independence between SNPs is evaluated using whole-genome forward simulations with recombination performed by the software tool MimicrEE (KOFLE and SCHLÖTTERER 2014). We generate a diploid founder population consisting of 2000 simulated haplotypes that mimic a wild population of *D. melanogaster* from Vienna (BASTIDE *et al.* 2013; KOFLE and SCHLÖTTERER 2014) and use the recombination rate of *D. melanogaster* (FISTON-LAVIER *et al.* 2010). Forward simulations are performed in 10 replicate runs. Allele counts are subjected to binomial sampling to mimic Pool-Seq with a given sequence coverage. N_e is estimated for non-overlapping 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 the performance of $N_e(P)$ to methods that correct only for a single sampling step. We assume binomial sampling out of an infinite gamete pool recommended by WAPLES (1989) and the weighting scheme of alleles put forward in JORDE and RYMAN (2007). Therefore we decided to compare $N_e(P)$ to the estimators proposed by these authors, and denote them by $N_e(W)$ and $N_e(JR)$, respectively.

We illustrate experimental sampling procedures by considering two major scenarios: (i) the full population is sequenced as one large pool; (ii) only a subset of the population is used to create pooled samples. Under scenario (i) we perform 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 ($S_j = N$) for our estimators. In this case, the coverage represents the sample size for estimators that correct only for a single sampling step. For scenario (ii), we sample individuals without replacement to generate pools of S_j individuals that are subjected to sequencing. Sequencing reads is modeled by binomial sampling as before. For this scenario again the coverage is taken as the sample size for $N_e(W)$ and $N_e(JR)$ estimators. Note, however, that the latter estimators permit correction only for a single sampling step, therefore the user has to decide on whether sampling individuals or reads is corrected for. In general, it is recommended to correct for the sampling step that contributes a larger extent to the sampling variance.

Genome-wide coverage differences are treated equally by subtracting correction defined by the corresponding coverage for each site before averaging over the number of SNPs for all methods.

Change point inference for the genome-wide estimates Selection causes change in allele frequency at a targeted site and also affects nearby sites through linkage. Consequently, deviation from the neutral expectation can result in a locally decreased N_e . To detect patterns of heterogeneous N_e along the genome we partition chromosomes into windows

of locally homogeneous N_e . To distinguish random fluctuations from systematic changes, we applied a segmentation algorithm, which guarantees type I error control, in the sense 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. The algorithm is implemented in the stepR software package (FRICK *et al.* 2014). It is related to a statistical multiscale change-point estimator (SMUCE) that has been suggested for partitioning sequences with respect to GC-content in (FUTSCHIK *et al.* 2014). As this method requires homogeneous variances, we applied it to log-transformed N_e estimates calculated genome-wide for non-overlapping windows. For visualization we back-transformed the obtained step function.

RESULTS AND DISCUSSION

Two-step correction is vital to avoid large bias in N_e estimates for long time series. Methods that do not correct for the additional sampling step, i.e. pooling, can lead to substantial bias in N_e estimates. We illustrate this (Fig. 2) on simulated data where the effective population size is inferred using allele frequency samples from the initial and evolved populations. 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). Figure 2 shows that the additional correction substantially decreases the bias for almost all data points. Methods that account only for a single sampling step may be severely biased. For the estimator $N_e(JR)$ the bias becomes less pronounced with an increasing number of generations. Under plan I, $N_e(P)$ is nearly unbiased, and plan II has a slight upward bias when applied on data simulated under plan I and the samples are taken at very close time points. 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.

[Figure 2 about here.]

Plan I and II estimators differ by a factor resulting from the covariance between the sample frequencies at generation 0 and t (equation (10)), which is inversely proportional

to the census population size. Consequently, the difference between plan 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 (Fig. S2-S4). 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 (Fig. S2). If, however, $N_e(P)$ 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 with more generations. When the entire population is sequenced as a single pool ($P = 100$), 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 $P = N$ 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 (Fig. S3 and S4). 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 plan I and II almost negligible (WAPLES 1989). Also, the sampling variance becomes proportionally smaller compared to the drift variance with 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 classical methods for both plans. In general, with Pool-Seq data the extent of the bias in $N_e(W)$ and $N_e(JR)$ estimates depends on the ratio between N and P , smaller pool sizes leading to a larger bias. Since we accounted for the sequencing step with these estimators (see section Estimating N_e on simulated data), decreasing the

coverage at a given pool size does not change the bias to a large extent 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 in subsequent simulations where $N_e = N$ is used but according to Fig. S3 and S4 $N_e(P)$ 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 Fig. 3. CV is defined as the ratio between the standard deviation and the mean ($CV = \sigma/\mu$, where both σ and μ 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 (Fig. 3 see also Fig. S6 and S8 where $N_e = 1000, t < 30, P \leq 100$ and $c = 50$). The bias is coming from a few outlier estimates, but the median shows more robust results (Fig. S10 and S11). For plan II estimators, the behavior of the method is similar (Fig. S5 and S7,S9). Note that the simulations underlying Fig. S5 and S7,S9 have been done under plan I sampling scheme.

[Figure 3 about here.]

Increasing the number SNPs reduces the variance of $N_e(P)$. We test how the number of loci 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 10000 (Fig. 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 the number of SNPs. Some further improvement is obtained when more than 10K SNPs are used (not shown), but the benefit of additional independent SNPs levels off. We conclude that $n = 1000$ independent 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.

[Figure 4 about here.]

Skew towards low frequency variants only moderately increases the variance of $N_e(\mathbf{P})$. In natural populations, the neutral site frequency spectrum is skewed towards low frequency variants. $N_e(P)$ 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 towards low frequency variants (exponential) as predicted 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 (Fig. 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 increasing number of generations the difference becomes negligible.

[Figure 5 about here.]

The presence of linkage disequilibrium does not severely bias the estimate. We investigated the sensitivity of our estimator to linkage disequilibrium between loci using genome-wide neutral simulations with recombination (KOFER and SCHLÖTTERER 2014). To estimate the effect of linkage disequilibrium, we simulated 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 non-overlapping windows with a fixed number of $n = 5000$ SNPs (Fig. 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 (Fig. 6).

[Figure 6 about here.]

Genome-wide heterogeneity of N_e is recovered in real data of *D. melanogaster*.

We analyzed data of a recent E&R study in *D. melanogaster* (OROZCO-TERWENGEL *et al.* 2012; TOBLER *et al.* 2014; FRANSSEN *et al.* 2015). In this experiment replicate populations of 1,000 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 using $N_e(P)$ under plan I. We considered non-overlapping windows of 10000 SNPs. To determine statistically significant changes in N_e along the genome we use the software provided by FRICK *et al.* (2014). This method requires homogeneity of variances. Since, the variance of estimates obtained with $N_e(P)$ 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 four biological replicates (Fig. 7).

The mean estimates for each chromosome arm as well as across the genome is stable across replicates (see table 1). As experimental evolution studies often aim to find signals that are consistent across replicates, this can be an important check of the experimental set up. 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.

All estimators mentioned above assume neutrality. Directional selection, however, leads to deterministic allele frequency changes that generate a larger deviation from the expected frequency than it would be due to random genetic drift alone. Consequently, estimates of N_e will be lower in the presence of directional selection. In the *D. melanogaster* example, N_e ranges from around 50 to almost 400. Around the centromere of chromosome 2, the estimated N_e decreases by two-third in replicates 1, 2 and 4. Even more striking, the estimated N_e is low on the entire chromosome arm 3R and also 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 segregating inversions (KAPUN *et al.* 2014) in combination with

selection potentially on rare variants. These results are consistent with TOBLER *et al.* (2014), who observed a massive amount of outlier SNPs around the centromere of chromosome 2 and on 3R. Strong LD, potentially due to selection on rare variants, could explain this inflated number of hitchhikers. Interestingly, certain regions of the genome show extensive differences in N_e between the replicates, which might be reflecting different selection histories possibly due to distinct genomic paths of adaptation.

[Figure 7 about here.]

[Table 1 about here.]

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 (Fig. 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. It is possible however to infer linkage information from the pooled sequence data using 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 is required, one can alternatively use a set of randomly distributed SNPs over the genome to obtain N_e estimate.

For 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 im-

portant component of an experimental design. For example, in repeatedly bottlenecked populations, the smallest population size dominates the N_e estimate (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 is available such changes can be detected by estimating N_e from pairwise comparisons between consecutive time points.

To estimate N_e we used theoretical expectations of the variance in allele frequency after t generations of neutral evolution. All evolutionary forces (selection, demography, etc.) that lead to deviations from the neutral expectations will affect our estimate. Nevertheless, systematic forces that result in a localized reduction in N_e , can be detected with the sliding window approach. The *Drosophila melanogaster* data set illustrates this point, i.e., the hypothesized region under selection exactly coincides with the region with reduced N_e (OROZCO-TERWENGEL *et al.* 2012; TOBLER *et al.* 2014; FRANSSEN *et al.* 2015).

Using a small number of generations can lead to outlier estimates. In general, $N_e(P)$ has a lower bias but larger variance, especially when t is small. We observe outlier estimates among replicates at early generations (generation 5, Fig. 2 and S2-S4) for $N_e(P)$. The deviation from the true N_e is especially large when the sampling variance is large compared to the drift variance ($N_e = 1000$, $P \leq 100$ and $c=50$, Fig. S8,S9). As pointed out by JORDE and RYMAN (2007) our weighting scheme leads to an increased variance but a smaller bias compared to other schemes. With this weighting design the signal for the N_e estimation is predominantly coming from the intermediate frequency variants. Having a large number of low frequency alleles thus has a similar effect as decreasing the number of SNPs.

To eliminate potential outliers and an inflated variance we recommend to increase the

signal to noise ratio by pooling 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 on the data, we suggest to use the genome-wide median taken across the estimates from individual windows because it is robust to outliers (Fig. 6). We also do not observe outliers when $N_e(P)$ is applied on the genome-wide data set of *Drosophila melanogaster* (table 1).

CONCLUSIONS

Effective population size 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 for this purpose and their performance was comprehensively evaluated on simulated as well as real data (GILBERT and WHITLOCK 2015; BARKER 2011; SERBEZOV *et al.* 2012; BAALSRUD *et al.* 2014; HOLLELEY *et al.* 2014). These studies mainly focused on genetic data collected from natural populations, which usually differs 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 over-dispersed, 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 well approximated by binomial sampling given that a large enough number of individuals is pooled. We do not model over-dispersion in our correction term for pooling, nevertheless, it would be possible to introduce a parameter that accounts for the additional between-pool variation causing over-dispersion.

We illustrate the applicability of our method for estimating N_e from experimental data of *Drosophila 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 χ^2 distribution (NEI and TAJIMA 1981) or alternatively apply non-parametric bootstrap approach and infer confidence limits based on sampling SNPs out of the original data set.

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

ACKNOWLEDGMENTS

We thank Mads Fristrup Schou, Susanne U. Franssen and Neda Barghi for helpful comments on the *Nest* software package. AJ and TT are part of the Vienna Graduate School of Population Genetics which is funded by the Austrian Science Fund (FWF, W1225). CS is also supported by the European Research Council grant "ArchAdapt", and TT is a recipient of a DOC Fellowship of the Austrian Academy of Sciences.

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, H. PRN, and H. JENSEN, 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, D. SCHNEIDER, R. E. LENSKE, and J. F. KIM, 2009, Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* *461*(7268): 1243–1247.
- BASTIDE, H., A. BETANCOURT, V. NOLTE, R. TOBLER, P. STÖBE, A. FUTSCHIK, and C. SCHLÖTTERER, 2013, A genome-wide, fine-scale map of natural pigmentation variation in *Drosophila melanogaster*. *PLoS Genet* *9*(6): e1003534.
- BOITARD, S., R. KOFLER, P. FRANÇOISE, D. ROBÉLIN, C. SCHLÖTTERER, and A. FUTSCHIK, 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.
- BOLLBACK, J. P., T. L. YORK, and R. NIELSEN, 2008, Estimation of $2N_e s$ from temporal allele frequency data. *Genetics* *179*(1): 497–502.
- BURKE, M. K., J. P. DUNHAM, P. SHAHRESTANI, K. R. THORNTON, M. R. ROSE, and A. D. LONG, 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.
- 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., 2009, Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet* *10*(3): 195–205.
- FALCONER, D. S. and T. F. C. MACKAY, 1996 *Introduction to Quantitative Genetics*. Benjamin Cummings.
- 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 Univ. 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, H. SHIM, A.-S. MALASPINAS, G. EWING, P. LIU, D. WEGMANN, D. R. CAFFREY, K. B. ZELDOVICH, D. N. BOLON, J. P. WANG, T. F. KOWALIK, C. A. SCHIFFER, R. W. FINBERG, and J. D. JENSEN, 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. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* *76*(3): 495–580.

FUTSCHIK, A., T. HOTZ, A. MUNK, and H. SIELING, 2014, Multiscale DNA partitioning: statistical evidence for segments. *Bioinformatics* **30**(16): 2255–2262.

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.

GAUTIER, M., J. FOUCAUD, K. GHARBI, T. CÉZARD, M. GALAN, A. LOISEAU, M. THOMSON, P. PUDLO, C. KERDELHUÉ, and A. ESTOUP, 2013, Estimation of population allele frequencies from next-generation sequencing data: pool-versus 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.

HILL, W. G., 1981 Estimation of effective population size from data on linkage disequilibrium. *Genetical Research* **38**: 209–216.

HOLLELEY, C. E., R. A. NICHOLS, M. R. WHITEHEAD, A. T. ADAMACK, M. R. GUNN, and W. B. SHERWIN, 2014 Testing single-sample estimators of effective population size in genetically structured populations. *Conservation Genetics* **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**(1): 285–293.

JORDE, P. E. and N. RYMAN, 2007, Unbiased estimator for genetic drift and effective population size. *Genetics* **177**(2): 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, and M. C. Whitlock, 2012, Experimental evolution. *Trends Ecol Evol* *27*(10): 547–560.

KIMURA, M., 1964 Diffusion model in population genetics. *J. App. Prob.*: 177–223.

KOFLE, R., P. OROZCO-TERWENGEL, N. DE MAIO, R. V. PANDEY, V. NOLTE, A. FUTSCHIK, C. KOSIOL, and C. SCHLÖTTERER, 2011 PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One* *6*(1): e15925.

KOFLE, R., R. V. PANDEY, and C. SCHLÖTTERER, 2011, PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* *27*(24): 3435–3436.

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

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*(1): 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 U S A* *105*(29): 10051–10056.

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. *Conservation Biology* *13*(3): 523–530.

MALASPINAS, A.-S., O. MALASPINAS, S. N. EVANS, and M. SLATKIN, 2012, Estimating allele age and selection coefficient from time-serial data. *Genetics* *192*(2): 599–607.

MATHIESON, I. and G. MCVEAN, 2013, Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. *Genetics* *193*(3): 973–984.

NEI, M. and F. TAJIMA, 1981, Genetic drift and estimation of effective population size. *Genetics* *98*(3): 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, and C. SCHLÖTTERER, 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*(4): 1055–1057.

POLLAK, E., 1983, A new method for estimating the effective population size from allele frequency changes. *Genetics* *104*(3): 531–548.

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*(1): 383–387.

REED, L. K., K. LEE, Z. ZHANG, L. RASHID, A. POE, B. HSIEH, N. DEIGHTON, N. GLASSBROOK, R. BODMER, and G. GIBSON, 2014, Systems genomics of metabolic phenotypes in wild-type *Drosophila melanogaster*. *Genetics* *197*(2): 781–793.

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.

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.

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*(2): 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, J. HERMISSON, and C. SCHLÖTTERER, 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. F., L. A. SALTER, and J. R. GOLD, 2001 Temporal-method estimates of N_e from highly polymorphic loci. *Conservation Genetics* 2: 297–308.

TURNER, T. L. and P. M. MILLER, 2012, Investigating natural variation in *Drosophila* courtship song by the evolve and resequence approach. *Genetics* 191(2): 633–642.

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.

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.

WANG, J. and M. C. WHITLOCK, 2003, Estimating effective population size and migration rates from genetic samples over space and time. *Genetics* 163(1): 429–446.

WAPLES, R. S., 1989, A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* 121(2): 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**(2): 633–644.

WILLIAMSON, E. G. and M. SLATKIN, 1999, Using maximum likelihood to estimate population size from temporal changes in allele frequencies. *Genetics* **152**(2): 755–761.

WRIGHT, S., 1931 Evolution in Mendelian populations. *Genetics* **16**(1-2): 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, Extra-binomial 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.

List of Figures

1	Two-step sampling in experimental evolution with <i>Drosophila</i>	33
2	Performance of different methods to estimate N_e based on simulated data . .	34
3	Coefficient of variation of $N_e(P)$ for plan I for various parameter values . . .	35
4	Effect of the number of SNPs used on the N_e estimates	36
5	Influence of the starting allele frequency distribution on $N_e(P)$ plan I estimator	37
6	Effect of linkage disequilibrium	38
7	Genome-wide N_e estimates from an E&R study in <i>D. melanogaster</i>	39

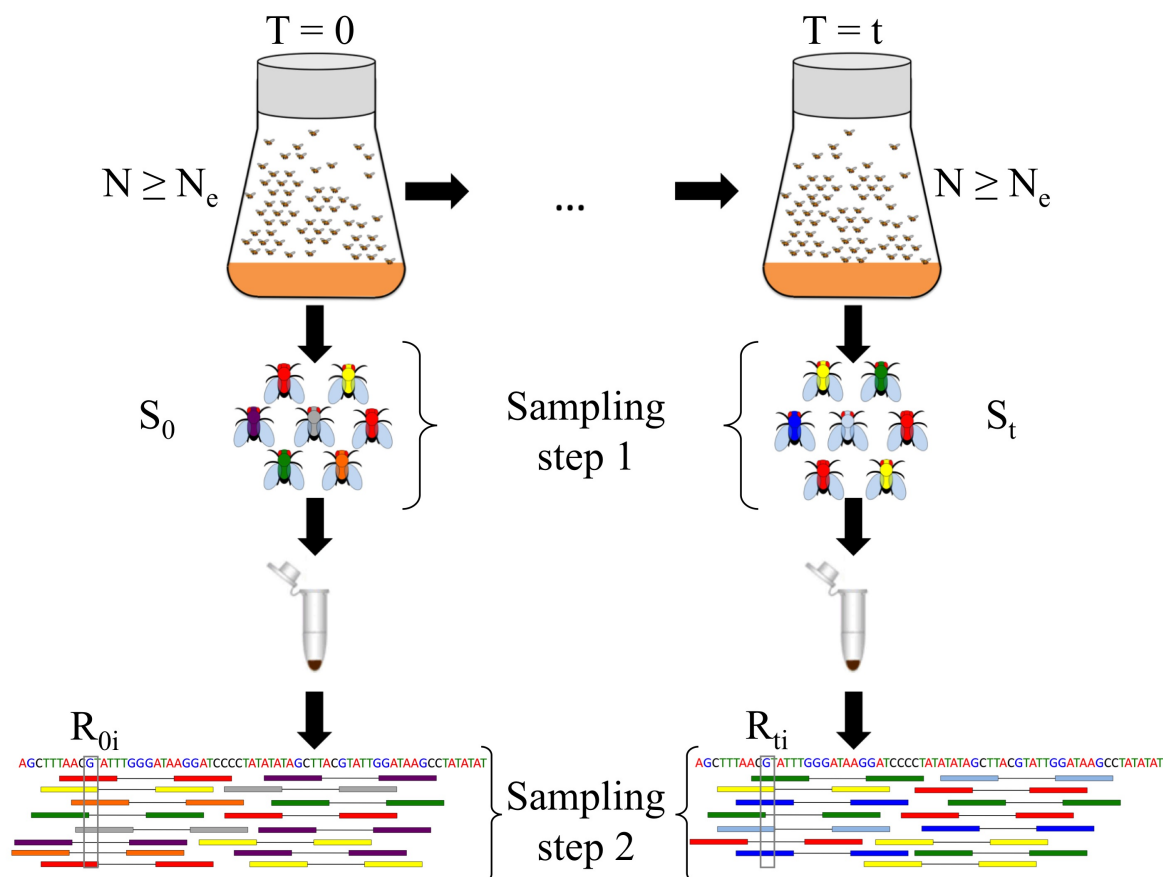


Figure 1: Two-step sampling in experimental evolution with *Drosophila*. In E&R studies, populations are propagated at a census size (N) defined by the experimenter, which is, in general, larger than the effective population size N_e . Using temporal methods N_e can be estimated from the variance in allele frequency in 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. We represent random variation in sequence coverage (R_j , $j \in \{0, t\}$) with 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 $T = j$) when correcting for the variance coming from sequencing.

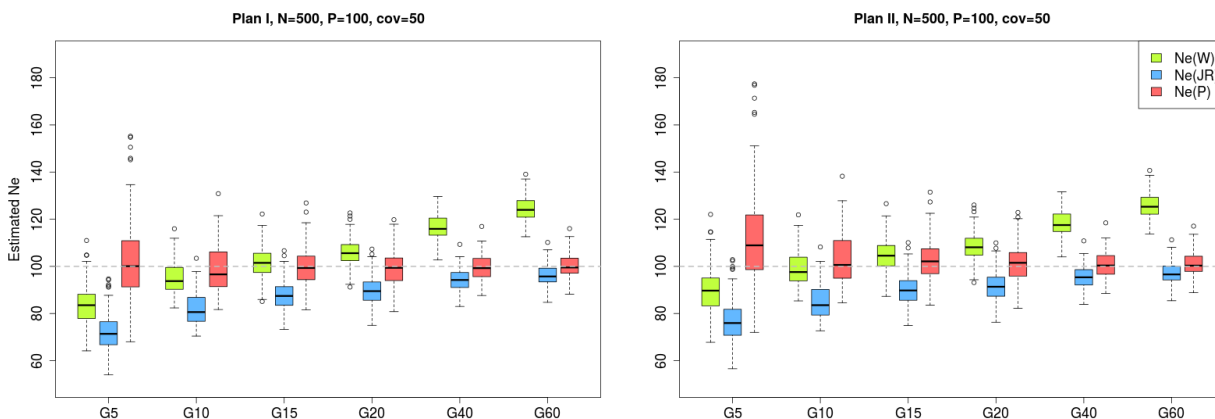


Figure 2: Performance of different methods to estimate N_e based on simulated data. Sixty generations of Wright-Fisher neutral evolution with $N_e = 100$ diploid individuals were simulated for 1000 unlinked loci (SNPs). Prior to sampling, the population was increased to a census size of $N = 500$ individuals at each generation. At the starting population (G0), and at each indicated time point a sample was taken to create a pool of $P=100$ individuals. Sequencing was modeled with an additional subsequent binomial sampling step of a random number of reads (50 on average) for each SNP. Then N_e was estimated on the resulting data set by separately contrasting G0 with each of the generations G5, G10, G15, G20, G40 and G60. We considered the estimators $N_e(W)$ (WAPLES 1989), $N_e(JR)$ (JORDE and RYMAN 2007) and our $N_e(P)$. Each box represents results from 100 simulations with identical parameters. The dashed grey line shows the true value of N_e . Data is simulated under plan I assumptions and the results of plan I and II estimators are shown in the left and right panel, respectively.

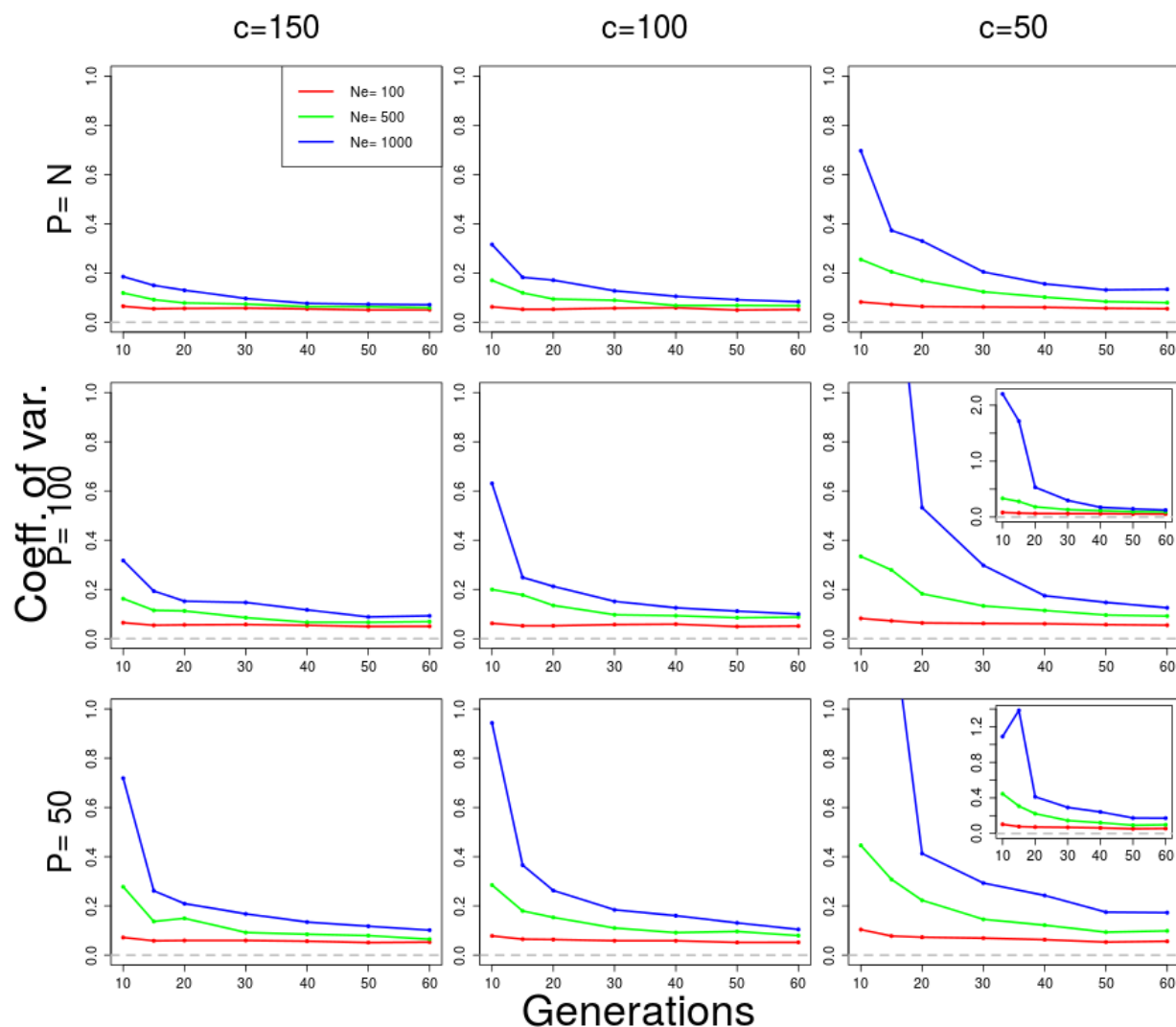


Figure 3: Coefficient of variation of $N_e(P)$ for plan I for various parameter values. Neutral Wright-Fisher simulations were performed with various combinations of the parameters effective population size ($N_e = 100, 500, 1000$ diploid individuals), pool size ($P = 100, 50$) and coverage ($c = 150, 100, 50$). $P = N$ indicates scenarios when the whole population is sequenced as a single pool. For all simulations, $N = N_e$ was used as census size of diploid individuals. Each value is calculated over 100 simulations. Negative estimates were excluded when calculating summary statistics. The frequency of negative estimates is listed in table S1. In situations where the coefficient of variation exceeds one, an inset figure shows the actual value.

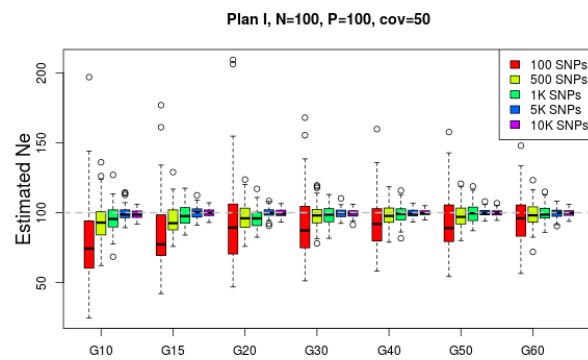


Figure 4: Effect of the number of SNPs used on the N_e estimates. The effective population size is estimated using $N_e(P)$ plan I on simulated data with $N_e = N = 100$. A total of $P = 100$ individuals is pooled and sequenced at a mean coverage of $c = 50$. Based on 100 simulation runs, N_e is estimated using different numbers of SNPs at multiple time points.

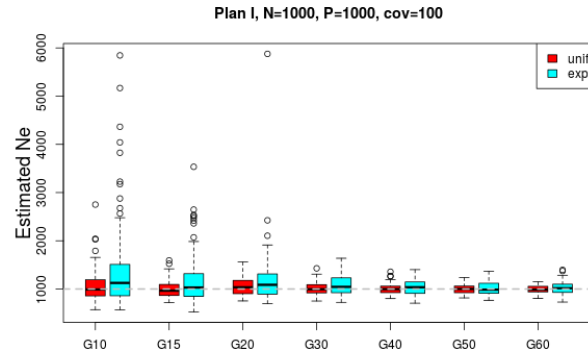


Figure 5: Influence of the starting allele frequency distribution on $N_e(P)$ plan I estimator. A comparison between uniform and exponentially distributed (neutral) starting allele frequencies is shown.

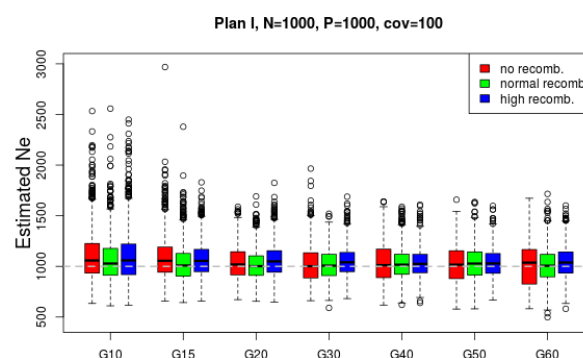


Figure 6: Effect of linkage disequilibrium. The effect of linkage disequilibrium on our estimators was evaluated based on a whole-genome forward simulation with recombination using the software MimicrEE (KOFER and SCHLÖTTERER 2014). Three sets of simulations were performed with different rates of recombination: high, normal and no recombination. For each parameter setup, a genome-wide simulation is replicated ten times. The effective population size was estimated with $N_e(P)$ plan I in non-overlapping windows of 5000 SNPs for each replicate. The distribution of N_e estimates across replicates and locations is shown in the box-plots.

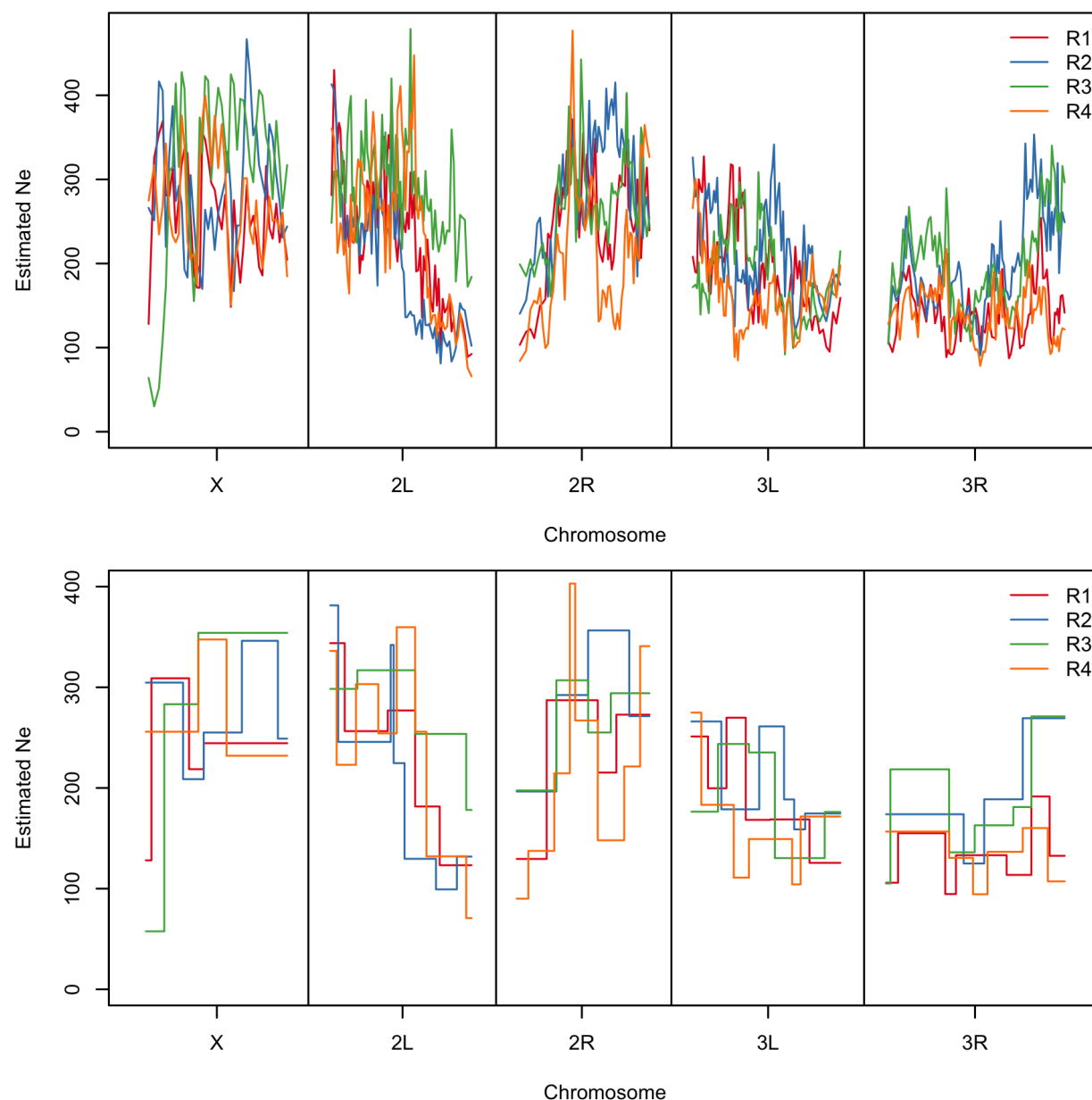


Figure 7: Genome-wide N_e estimates from an E&R study in *D. melanogaster*. N_e is estimated based on the allele frequency changes between founder and evolved populations at generation 59 (FRANSEN *et al.* 2015). In the top panel genome-wide estimates calculated with $N_e(P)$ plan I using non-overlapping windows of 10000 SNPs are shown. Chromosome-wide mean estimates are shown in table 1. Statistically significant changes in N_e is determined after log-transformation of the data (bottom panel). Chromosome arms are separated by vertical bars. Interestingly, the estimates on the X chromosome do not reflect the expectation of N_e values reduced by three-quarters compared to autosomes.

List of Tables

1	Genome-wide mean and median N_e estimates from an E&R study in <i>D. melanogaster</i>	40
---	---	----

	Mean						Median					
	X	2L	2R	3L	3R	Genome-wide	X	2L	2R	3L	3R	Genome-wide
R1	257	235	250	193	139	200	256	232	253	186	139	191
R2	284	202	293	215	200	223	274	221	292	207	190	216
R3	312	293	270	192	203	237	333	291	269	191	191	237
R4	266	243	216	165	138	187	252	256	207	160	136	164

Table 1: Genome-wide mean and median N_e estimates from an E&R study in *D. melanogaster*. The effective population size is estimated with $N_e(P)$ plan I in windows of 10000 SNPs (Fig. 7). The mean and median estimates across windows are shown for the major chromosome arms. Genome-wide mean and median is taken over the autosomes excluding chromosome 4.